Handbook of Paleoanthropology
We dedicate these volumes to our long-time colleagues Hartmut Rothe and Theodoros Pitsios in appreciation of their friendship and unique contributions to primatology and paleoanthropology.
Preface to the Series

Palaeoanthropology is perhaps the most multidisciplinary of all the sciences. Any complete account of the evolution and of the cultural and biological contexts of *Homo sapiens* must combine information from geology, paleoecology, primatology, evolutionary biology and a host of other fields. Above all, historical information garnered from the fossil record needs to be combined with, and interpreted in the light of, what we know of the living world. In these volumes we have brought together contributions by a variety of leading specialists that reflect the broad spectrum of modern paleoanthropology, in an attempt to provide a resource that we hope will be useful to professionals and students alike.

Volume I of this three-volume Handbook deals with principles, methods, and approaches. In recent years enormous advances have been made in such areas as phylogenetic analysis, evolutionary theory and philosophy, paleoecology, and dating methods. The contributions aim to present the state of the art in these and other relevant fields, as well as to furnish succinct introductions to them and to reflect the many ways in which they interact. Human beings are primates, and Volume II is devoted to primate origins, evolution, behavior, and adaptive variety. In this compilation the emphasis is on the integration of fossil data with the vast amount that is now known of the behavior and ecology of living primates in natural environments. The third and final volume deals directly with the fossil and molecular evidence for the evolution of *Homo sapiens* and its fossil relatives (the family Hominidae or subfamily Homininae, according to taste, a matter that we have left to each individual contributor). Paleoanthropology is a pluralistic and actively developing field in which much remains to be settled, and we have not tried to impose any uniformity of viewpoint on our authors. Instead, while maintaining an emphasis on the data, we have encouraged them to express their individual interpretations rather than to cover all possible points of view. This has inevitably led to a certain degree of heterogeneity of opinion between the covers of this Handbook; but we believe that this is the best way of reflecting the excitement and momentum of the field and that it is best for the reader to be left to reach his or her own conclusions. Science is, after all, a process rather than a static product, and one of our primary aims here is to reflect the ongoing dynamism of that process in paleoanthropology.

We thank all of the contributors to these volumes for their participation. Some initially responded enthusiastically while others needed convincing about the basic strategy of the Handbook, but all responded marvellously to the
particular needs of a corporate effort such as this one. We are particularly grateful to those authors who responded at short notice to needs that became apparent only as the project progressed. This series was conceived in collaboration with Prof. Hartmut Rothe of the University of Göttingen, who was later forced to withdraw for reasons beyond his control. We thank him most warmly for his creativity in the conceptual stages and for his subsequent moral support. The laborious process of putting together the volumes could not have been accomplished without the cheerful help of Thorolf Hardt, whose active involvement was indispensable throughout.

This project could never have come to fruition without the enthusiastic support of Dr. Dieter Czeschlik, editor life sciences at Springer Publishing, and the efficient assistance of Mrs. Ursula Gramm. We express our deep gratitude to Mrs. Susanne Friedrichsen and Mrs. Caroline Simpson, who showed both care and commitment during the phases of copyediting and product development. The continuous cooperation and dialogue with them and their professionalism gave us the courage to see the project through. Further thanks go to Mrs. Sandra Fabiani and her colleagues at Springer Publishing, who prepared the eReference. Ms. Nitya Swaruba, compositor at SPi Technologies, deserves warm thanks for her efficient help, and finally our gratitude goes in addition to Britta Hardt, Peter Menke and Monika Sandführ, who also rendered much valuable assistance.

Winfried Henke and Ian Tattersall
Mainz and New York City
November, 2006
Table of Contents

Volume 1

1 Historical Overview of Paleoanthropological Research ............... 1
   Winfried Henke

2 Evolutionary Theory in Philosophical Focus ....................... 57
   Philippe Huneman

3 The Ontogeny–Phylogeny Nexus in a Nutshell: Implications for
   Primatology and Paleoanthropology ............................ 103
   Peter R. Menke

4 Principles of Taxonomy and Classification: Current Procedures
   for Naming and Classifying Organisms .......................... 141
   Michael Ohl

5 Quantitative Approaches to Phylogenetics .......................... 167
   Kaila E. Folinsbee · David C. Evans · Jörg Fröbisch · Linda A. Tsuji ·
   Daniel R. Brooks

6 Homology: A Philosophical and Biological Perspective ............. 217
   Olivier Rieppel

7 Taphonomic and Diagenetic Processes .............................. 241
   Gisela Grupe

8 Archeology .................................................. 261
   Miriam N. Haidle

9 Contribution of Stable Light Isotopes to Paleoenvironmental
   Reconstruction .................................................. 289
   Julia Lee-Thorp · Matt Sponheimer

10 Chronometric Methods in Paleoanthropology ..................... 311
    Günther A. Wagner
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>Geological Background of Early Hominid Sites in Africa</td>
<td>339</td>
</tr>
<tr>
<td></td>
<td>Ottmar Kullmer</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Paleoclimate</td>
<td>357</td>
</tr>
<tr>
<td></td>
<td>Keith Alverson</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Paleosols</td>
<td>383</td>
</tr>
<tr>
<td></td>
<td>Gregory Retallack</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Quaternary Deposits and Paleosites</td>
<td>409</td>
</tr>
<tr>
<td></td>
<td>Klaus-Dieter Jäger</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Zoogeography: Primate and Early Hominin Distribution and Migration</td>
<td>421</td>
</tr>
<tr>
<td></td>
<td>Patterns</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Alan Turner · Hannah O’Regan</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Patterns of Diversification and Extinction</td>
<td>441</td>
</tr>
<tr>
<td></td>
<td>Walter Etter</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Paleoecology: An Adequate Window on the Past?</td>
<td>503</td>
</tr>
<tr>
<td></td>
<td>Thorolf Hardt · Britta Hardt · Peter R. Menke</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Hominin Paleodiets: The Contribution of Stable Isotopes</td>
<td>555</td>
</tr>
<tr>
<td></td>
<td>Matt Sponheimer · Julia Lee · Thorp</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Estimation of Basic Life History Data of Fossil Hominoids</td>
<td>587</td>
</tr>
<tr>
<td></td>
<td>Helmut Hemmer</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Population Genetics and Paleoanthropology</td>
<td>621</td>
</tr>
<tr>
<td></td>
<td>John H. Relethford</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>Ancient DNA</td>
<td>643</td>
</tr>
<tr>
<td></td>
<td>Susanne Hummel</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>Paleodemography of Extinct Hominin Populations</td>
<td>673</td>
</tr>
<tr>
<td></td>
<td>Janet Monge · Alan Mann</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Modeling the Past: The Primatological Approach</td>
<td>701</td>
</tr>
<tr>
<td></td>
<td>R. W. Sussman · Donna Hart</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>Modeling the Past: The Paleoethnological Evidence</td>
<td>723</td>
</tr>
<tr>
<td></td>
<td>Paolo Biagi</td>
<td></td>
</tr>
</tbody>
</table>
## Volume 2

1. **Primate Origins and Supraordinal Relationships: Morphological Evidence** ................................................... 831  
   Mary T. Silcox · Eric J. Sargis · Jonathan I. Bloch · Doug M. Boyer

2. **Molecular Evidence on Primate Origins and Evolution** ............................................................. 861  
   Hans Zischler

3. **Fossil Record of the Primates from the Paleocene to the Oligocene** ................................................. 889  
   D. Tab Rasmussen

4. **Fossil Record of Miocene Hominoids** ......................... 921
   David R. Begun

5. **The Biotic Environments of the Late Miocene Hominids** ......................................................... 979  
   Jordi Agustí

6. **Postcranial and Locomotor Adaptations of Hominoids** ......................................................... 1011  
   Carol V. Ward

7. **Hominoid Cranial Diversity and Adaptation** ................. 1031  
   Alan Bilsborough · Todd C. Rae

8. **Dental Adaptations of African Apes** ............................................. 1107  
   Mark F. Teaford · Peter S. Ungar

9. **Evolution of the Primate Brain** ............................................. 1133  
   Dean Falk
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>Primate Life Histories</td>
<td>1163</td>
</tr>
<tr>
<td></td>
<td>Elke Zimmermann · Ute Radespiel</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>The Biology and Evolution of Ape and Monkey Feeding</td>
<td>1207</td>
</tr>
<tr>
<td></td>
<td>Joanna E. Lambert</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Great Ape Social Systems</td>
<td>1235</td>
</tr>
<tr>
<td></td>
<td>Angela Meder</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Primate Intelligence</td>
<td>1273</td>
</tr>
<tr>
<td></td>
<td>Richard W. Byrne</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Chimpanzee Hunting Behavior</td>
<td>1295</td>
</tr>
<tr>
<td></td>
<td>Nicholas E. Newton-Fisher</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Cooperation, Coalition, and Alliances</td>
<td>1321</td>
</tr>
<tr>
<td></td>
<td>Charlotte K. Hemelrijk · Jutta Steinhauser</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Volume 3</strong></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Potential Hominoid Ancestors for Hominidae</td>
<td>1347</td>
</tr>
<tr>
<td></td>
<td>George D. Koufos</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Defining Hominida</td>
<td>1379</td>
</tr>
<tr>
<td></td>
<td>Jeffrey H. Schwartz</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Origins of Homininae and Putative Selection Pressures Acting on the Early Hominins</td>
<td>1409</td>
</tr>
<tr>
<td></td>
<td>Boguslaw Pawłowski</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Role of Environmental Stimuli in Hominid Origins</td>
<td>1441</td>
</tr>
<tr>
<td></td>
<td>Elisabeth S. Vrba</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>The Origins of Bipedal Locomotion</td>
<td>1483</td>
</tr>
<tr>
<td></td>
<td>William E. H. Harcourt-Smith</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>The Earliest Putative Hominids</td>
<td>1519</td>
</tr>
<tr>
<td></td>
<td>Brigitte Senut</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>The Species and Diversity of Australopiths</td>
<td>1539</td>
</tr>
<tr>
<td></td>
<td>William H. Kimbel</td>
<td></td>
</tr>
<tr>
<td>Chapter</td>
<td>Title</td>
<td>Pages</td>
</tr>
<tr>
<td>---------</td>
<td>-------------------------------------------------------------------------------------------------</td>
<td>--------</td>
</tr>
<tr>
<td>8</td>
<td>Defining the Genus <em>Homo</em></td>
<td>1575</td>
</tr>
<tr>
<td></td>
<td>Mark Collard · Bernard Wood</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>The Earliest Putative <em>Homo</em> Fossils</td>
<td>1611</td>
</tr>
<tr>
<td></td>
<td>Friedemann Schrenk · Ottmar Kullmer · Timothy Bromage</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td><em>Homo ergaster</em> and Its Contemporaries</td>
<td>1633</td>
</tr>
<tr>
<td></td>
<td>Ian Tattersall</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Defining <em>Homo erectus</em>: Size Considered</td>
<td>1655</td>
</tr>
<tr>
<td></td>
<td>Susan C. Antón · Fred Spoor · Connie D. Fellmann · Carl C. Swisher III</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Later Middle Pleistocene <em>Homo</em></td>
<td>1695</td>
</tr>
<tr>
<td></td>
<td>G. Philip Rightmire</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Neanderthals and Their Contemporaries</td>
<td>1717</td>
</tr>
<tr>
<td></td>
<td>Katerina Harvati</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Origin of Modern Humans</td>
<td>1749</td>
</tr>
<tr>
<td></td>
<td>Günter Bräuer</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Analyzing Hominid Phylogeny</td>
<td>1781</td>
</tr>
<tr>
<td></td>
<td>David Strait · Frederick E. Grine · John G. Fleagle</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Phylogenetic Relationships (Biomolecules)</td>
<td>1807</td>
</tr>
<tr>
<td></td>
<td>Todd R. Disotell</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Population Biology and Population Genetics of Pleistocene Hominins</td>
<td>1825</td>
</tr>
<tr>
<td></td>
<td>Alan R. Templeton</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Species Concepts and Speciation: Facts and Fantasies</td>
<td>1861</td>
</tr>
<tr>
<td></td>
<td>Colin Groves</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Human Environmental Impact in the Paleolithic and Neolithic</td>
<td>1881</td>
</tr>
<tr>
<td></td>
<td>Wolfgang Nentwig</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>The Dentition of American Indians: Evolutionary Results and Demographic Implications Following Colonization from Siberia</td>
<td>1901</td>
</tr>
<tr>
<td></td>
<td>Christy G. Turner II · G. Richard Scott</td>
<td></td>
</tr>
<tr>
<td>Chapter</td>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>21</td>
<td>Overview of Paleolithic Archeology</td>
<td>1943</td>
</tr>
<tr>
<td></td>
<td><em>Nicholas Toth · Kathy Schick</em></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>The Network of Brain, Body, Language, and Culture</td>
<td>1965</td>
</tr>
<tr>
<td></td>
<td><em>Steven Mithen</em></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Cultural Evolution in Africa and Eurasia During the Middle and Late Pleistocene</td>
<td>2001</td>
</tr>
<tr>
<td></td>
<td><em>Nicholas Conard</em></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>Paleoanthropology and the Foundation of Ethics: Methodological Remarks on the Problem of Criteriology</td>
<td>2039</td>
</tr>
<tr>
<td></td>
<td><em>Mathias Gutmann · Michael Weingarten</em></td>
<td></td>
</tr>
</tbody>
</table>
About the Editors

Editors in Chief

Prof. Dr. Dr. h. c. Winfried Henke
Institut für Anthropologie (1050)
Fachbereich 10 - Biologie
Johannes Gutenberg-Universität Mainz
D-55099 Mainz
Germany
email: henkew@uni-mainz.de

Prof. Dr. Ian Tattersall
Division of Anthropology
American Museum of Natural History
New York, NY 10024–5192
USA
email: iant@amnh.org

Editorial Assistant

Dipl. Biol. Thorolf Hardt
Institut für Anthropologie (1050)
Fachbereich 10 - Biologie
Johannes Gutenberg-Universität Mainz
D-55099 Mainz
Germany
email: thormuel@students.uni-mainz.de

Winfried Henke is currently Professor of Anthropology at the Johannes Gutenberg University of Mainz. He was born in 1944 in Ludwigshorst/Pomerania, Germany, and studied biology, anthropology and geosciences in Kiel and Braunschweig. He received his Ph.D. from the University of Kiel in 1971, his thesis focusing on a prehistoric anthropological topic. In 1990 he habilitated with a monograph on the “Anthropology of Early Paleolithics and Mesolithics” at the University of Mainz.

Research activities in various countries (Iceland, Israel, Jordan, US, Greece) and extensive teaching assignments in the Erasmus exchange program at numerous European Universities followed. From 1996 to 2004, he acted as anthropology referee for the German Research Foundation (DFG). He served at the advisory boards of many scientific journals, and was advisory consultant to museums, e.g. the Neanderthal Museum (Mettmann, Germany). In 2006 he was awarded with the honorary doctorate of the National and Kapodistrian University of Athens and is an elected member of the German Academy of
Sciences Leopoldina. Areas of research and teaching: paleoanthropology, primatology, prehistoric anthropology, comparative morphology, systematics, demography and sociobiology.

He has published approximately 180 original papers in scientific journals and anthologies, and over 600 book reviews. He is author, co-author (together with H. Rothe) and editor of several books, including such standard works as “Paläoanthropologie” and “Stammesgeschichte des Menschen” (published at Springer-Verlag).

**Ian Tattersall** is currently Curator in the Division of Anthropology of the American Museum of Natural History in New York City. Born in England and raised in East Africa, he has carried out both primatological and paleontological fieldwork in countries as diverse as Madagascar, Vietnam, Surinam, Yemen and Mauritius. Trained in archeology and anthropology at Cambridge, and in geology and vertebrate paleontology at Yale, Tattersall has concentrated his research since the 1960s in two main areas: the analysis of the human fossil record and its integration with evolutionary theory, and the study of the ecology and systematics of the lemurs of Madagascar. Tattersall is also a prominent interpreter of human paleontology to the public, with several trade books to his credit, among them *The Monkey in the Mirror* (2002), *Extinct Humans* (with Jeffrey Schwartz, 2000), *Becoming Human: Evolution and Human Uniqueness* (1998) and *The Last Neanderthal: The Rise, Success and Mysterious Extinction of Our Closest Human Relatives* (1995; rev. 1999) as well as several articles in *Scientific American* and the co-editorship of the definitive *Encyclopedia of Human Evolution and Prehistory*. He lectures widely, and, as curator, has also been responsible for several major exhibits at the American Museum of Natural History, including *Ancestors: Four Million Years of Humanity* (1984); *Dark Caves, Bright Visions: Life In Ice Age Europe* (1986); *Madagascar: Island of the Ancestors* (1989); *The First Europeans: Treasures from the Hills of Atapuerca* (2003); and the highly acclaimed *Hall of Human Biology and Evolution* (1993).

**Thorolf Hardt** graduated in biology and is currently a PhD student at the Institute of Anthropology, University of Mainz. He was born in 1973 in Neustadt (RhP/Germany) and studied anthropology, paleontology and zoology in Kiel and Mainz. At present his research activities are focused on Geometric Morphometrics, functional morphology and evolutionary biology in Primates.
Contributors

J. Agustí Ballester
ICREA-Institut of Human Paleoenecology,
Universitat Rovira i Virgili,
Pl. Imperial Tarraco, 1,
43005-Tarragona,
Spain

K. Alverson
Ocean Observations and Services,
IOC/UNESCO,
Global Ocean Observing System,
1 rue Miollis,
75732 Paris Cedex 15,
France

S. C. Antón
Center for the Study of Human Origins,
Department of Anthropology NYU,
25 Waverly Place,
New York, NY 10003,
USA

D. R. Begun
Department of Anthropology,
100 St George Street, Rm 1037,
University of Toronto,
Toronto, ONT M5S 3G3,
Canada

D. M. Boyer
Department of Anatomical Sciences,
Stony Brook University,
Stony Brook, NY 11794-8081,
USA

A. Bilsborough
Department of Anthropology,
University of Durham,
43 Old Elvet, Durham,
DH 1 3HN,
UK

J. I. Bloch
Vertebrate Paleontology,
Florida Museum of Natural History,
Dickenson Hall, University of Florida,
Gainesville, Florida 32611-7800,
USA

P. Biagi
Department of Science of Antiquities and the Near East,
Ca’ Foscari University, Venice,
Palazzo Bernardo, S. Polo 1977,
30125 Venezia,
Italy

G. Bräuer
Institut für Humanbiologie,
Universität Hamburg,
Allende-Platz 2,
D-20146 Hamburg,
Germany
T. Bromage  
Department of Biomaterials and Biomimetics,  
New York University College of Dentistry,  
345 East 24th Street, Room 804-S,  
New York, NY 10010, USA

D. R. Brooks  
Department of Ecology & Evolutionary Biology,  
University of Toronto,  
25 Harbord Street,  
Toronto, Ontario M5S 3G5, Canada

R. W. Byrne  
School of Psychology,  
University of St. Andrews,  
St. Andrews, Fife KY16 9JU, Scotland

M. Collard  
Laboratory of Biological Anthropology,  
Department of Anthropology,  
University of British Columbia,  
6303 NW Marine Drive,  
Vancouver, British Columbia V6T 1Z1, Canada

B. Comrie  
Department of Linguistics,  
Max Planck Institute of Evolutionary Anthropology,  
Deutscher Platz 6,  
D-04103 Leipzig, Germany

N. J. Conard  
Institut für Ur- und Frühgeschichte,  
Abteilung Ältere Urgeschichte und Quartärökologie,  
Eberhard-Karls-Universität Tübingen,  
Schloss Hohentübingen,  
D-72070 Tübingen, Germany

T. R. Disotell  
Center for the Study of Human Origins,  
Department of Anthropology,  
New York University,  
25 Waverly Place,  
New York, NY 10003, USA

W. Etter  
Naturhistorisches Museum Basel,  
Abteilung Geowissenschaften,  
Augustinergasse 2,  
CH-4001 Basel, Switzerland

D. Evans  
Department of Biology,  
University of Toronto at Mississauga,  
3359 Mississauga Road,  
Mississauga, ON L5L 1C6, Canada

D. Falk  
Department of Anthropology,  
Florida State University,  
Tallahassee,  
FL 32306-7772, 1847 West, Tennessee St., USA
C. D. Fellmann  
Center for the Study of Human Origins,  
Department of Anthropology,  
New York University,  
25 Waverly Place,  
New York City, NY 10003,  
USA

J. G. Fleagle  
Department of Anatomical Sciences,  
Health Sciences Centre, Stony Brook University,  
Stony Brook, New York 11794-8081,  
USA

K. E. Folinsbee  
Department of Biology,  
University of Toronto at Mississauga,  
3359 Mississauga Road,  
Mississauga, ON L5L 1C6,  
Canada

J. Fröbisch  
Department of Biology,  
University of Toronto at Mississauga,  
3359 Mississauga Road North,  
Mississauga, ON L5L 1C6,  
Canada

F. E. Grine  
Department of Anthropology,  
State University of New York,  
Stony Brook, NY 11794-4364,  
USA

C. P. Groves  
School of Archaeology and Anthropology,  
Australian National University

Canberra, ACT 0200,  
Australia

G. Grupe  
Dept. I für Biologie,  
Bereich Biodiversitätsforschung/Anthropologie,  
Universität München,  
Grosshaderner Straße 2,  
D-82152 Planegg-Martinsried,  
Germany

M. Gutmann  
Institut für Philosophie,  
Philipps-Universität Marburg,  
Wilhelm-Röpke-Strasse 6, Block B  
D-35032 Marburg,  
Germany

M. N. Haidle  
Institut für Ur- und Frühgeschichte und Archäologie des Mittelalters,  
Abt. Ältere Urgeschichte und Quartärökologie,  
Universität Tübingen,  
Burgsteige 11,  
D-72070 Tübingen,  
Germany

W. E. H. Harcourt-Smith  
Division of Paleontology,  
American Museum of Natural History,  
Central Park West and 79th Street,  
New York, NY 10024,  
USA

B. Hardt  
Institut für Anthropologie (1050),  
Fachbereich 10 – Biologie,  
Johannes Gutenberg-Universität
Mainz,
D-55099 Mainz,
Germany

T. Hardt
Institut für Anthropologie (1050),
Fachbereich 10 – Biologie,
Johannes Gutenberg-Universität
Mainz,
D-55099 Mainz,
Germany

D. L. Hart
Department of Anthropology,
University of Missouri at St. Louis,
St. Louis, MO 63130, USA

K. Harvati
Department of Human Evolution,
Max Planck Institute for
Evolutionary Anthropology,
Deutscher Platz 6,
D-04103 Leipzig,
Germany

C. K. Hemelrijk
Theoretical Biology,
Centre for Ecological and
Evolutionary Studies,
University of Groningen,
Biological Centre,
Kerklaan 30,
9751 NN Haren,
The Netherlands

H. Hemmer
Anemonenweg 18,
D-55129 Mainz,
Germany

or

Institut für Zoologie,
Fachbereich 10 – Biologie,
Johannes Gutenberg-Universität
Mainz,
Johannes v. Müller-Weg 6,
D-55128 Mainz,
Germany

W. Henke
Institut für Anthropologie (1050),
Fachbereich 10 – Biologie,
Johannes Gutenberg-Universität
Mainz,
D-55099 Mainz,
Germany

J.-J. Hublin
Department of Human Evolution,
Max Planck Institute for Evolutionary
Anthropology,
Deutscher Platz 6,
D-4103 Leipzig,
Germany

S. Hummel
Johann Friedrich Blumenbach-
Institut für Zoologie und
Anthropologie, Historische
Anthropologie und Humanökologie,
Georg August-Universität Göttingen,
Bürgerstrasse 50,
D-37037 Göttingen,
Germany

P. Huneman
Institut d’Histoire et de Philosophie
des Sciences et des Techniques
CNRS/Université Paris I
Sorbonne,
K.-D. Jäger
Institut für Prähistorische Archäologie,
Martin Luther-Universität Halle,
Brandbergweg 23c,
D-06099 Halle,
Germany

W. H. Kimbel
Institute of Human Origins,
Arizona State University,
P.O. Box 874101,
Tempe, AZ 85287-4101,
USA

G. D. Koufos
Department of Geology,
University of Thessaloniki,
GR-54124 Thessaloniki,
Greece

O. Kullmer
Forschungsinstitut und Naturmuseum,
Abt. Paläoanthropologie und Quartärpaläontologie,
Senckenberganlage 25,
D-60325 Frankfurt am Main,
Germany

J. E. Lambert
Departments of Anthropology and Zoology,
University of Wisconsin-Madison,
5317 Wm Sewell Social Science Building,
Madison, WI 53706,
USA

J. Lee-Thorp
Department of Archaeological Sciences,
University of Bradford,
Bradford, West Yorkshire,
BD7 1DP,
UK

G. A. Macho
Palaeoanthropology Research Group,
Centre for Research of Evolutionary Anthropology,
Whitelands College,
Roehampton University,
London SW15 4JD,
England

A. E. Mann
Department of Anthropology,
Princeton University,
Princeton,
NJ 08544,
USA

A. Meder
Augustenstrasse 122,
D-70197 Stuttgart,
Germany

P. Menke
Institut für Anthropologie (1050),
Fachbereich 10 – Biologie,
Johannes Gutenberg-Universität Mainz,
D-55099 Mainz,
Germany
S. Mithen
School of Human and Environmental Sciences,
The University of Reading,
Whiteknights, P.O. Box 217,
Reading, Berkshire,
RG6 6AHm UK

J. Monge
Department of Anthropology,
University of Pennsylvania,
University Museum,
3260 South Street,
Philadelphia, PA 19104-6398,
USA

W. Nentwig
Zoologisches Institut,
Universität Bern,
Baltzer Strasse 6,
CH-3012 Bern,
Switzerland

N. E. Newton-Fisher
Department of Anthropology,
Marlowe Building,
University of Kent,
Canterbury, CT2 7NR,
UK

H. O’Regan
School of Biological and Earth Sciences,
Liverpool John Moores University,
Liverpool L3 3AF,
UK

M. Ohl
Humboldt-Universität zu Berlin,
Museum für Naturkunde,
D-10115 Berlin,
Invalidenstraße 43,
Germany

B. Pawłowski
Department of Anthropology,
University of Wrocław,
ul. Kuznica 35,
Wrocław 50-138,
Poland

U. Radespiel
Institut für Zoologie,
Tierärztliche Hochschule Hannover,
Bünteweg 17,
D-30559 Hannover,
Germany

T. C. Rae
Department of Anthropology,
University of Durham,
43 Old Elvet, Durham,
DH1 3HN,
UK

D. T. Rasmussen
Department of Anthropology,
CB 1114, Washington University,
One Brookings Drive,
St. Louis, Missouri 63130,
USA

J. H. Relethford
Department of Anthropology,
SUNY College at Oneonta,
Oneonta, NY 13820,
USA

G. J. Retallack
Department of Geology,
University of Oregon,
Eugene, OR 97403-1272, USA

**O. Rieppel**
Department of Geology, The Field Museum, 1400 S. Lake Shore Drive, Chicago, IL 60605-2496, USA

**G. P. Rightmire**
Department of Anthropology, Peabody Museum, Harvard University, Cambridge, MA 02138, USA

**E. J. Sargis**
Department of Anthropology, Yale University, P.O. Box 208277, New Haven, CT 06520, USA

**K. Schick**
Stone Age Institute, 1392 W. Dittemore Road, Gosport, IN 47433, USA

**F. Schrenk**
Forschungsinstitut und Naturmuseum, Abt. Palaeanthropologie und Quartärlaentologie, Senckenberanganlage 25, D-60325 Frankfurt am Main, Germany

J. H. Schwartz
Department of Anthropology, 3302 WWPH, University of Pittsburgh, Pittsburgh, PA 15260, USA

**G. R. Scott**
Department of Anthropology, University of Nevada, Reno, NV 89557, USA

**B. Senut**
Laboratoire de Paléontologie, Museum National d’Histoire Naturelle & UMR 8569, CNRS, 8, rue Buffon, 75005 Paris, France

**M. T. Silcox**
University of Winnipeg, Department of Anthropology, 515 Portage Ave., Winnipeg, Manitoba R3B 2E9, Canada

**M. Sponheimer**
Department of Anthropology, University of Colorado at Boulder, Boulder, CO 80309, USA

**F. Spoor**
Department of Anatomy & Developmental Biology,
University College London, Gower St., London WC1E 6BT, UK

J. Steinhauser
Theoretical Biology Group, Centre for Ecological and Evolutionary Studies, University of Groningen, P.O. Box 14, 9750 AA Haren, The Netherlands

D. Strait
Department of Anthropology, University at Albany, 1400 Washington Avenue, Albany, New York 12222, USA

R. W. Sussman
Department of Anthropology, Washington University at St. Louis, St. Louis, MO 63130, USA

C. C. Swisher III
Department of Geological Sciences, Rutgers University, 610 Taylor Road, Piscataway, USA

I. Tattersall
Division of Anthropology, American Museum of Natural History, New York, NY 10024-5192, USA

M. F. Teaford
Center for Functional Anatomy and Evolution, Johns Hopkins University School of Medicine, 1830 E. Monument St., Room 303, Baltimore, MD 21205, USA

A. R. Templeton
Department of Biology, Washington University in St. Louis, 1 Brookings, Campus Box 1137, St. Louis, MO 63130, USA

N. Toth
Stone Age Institute, 1392 W. Dittemore Road, Gosport, IN 47433, USA

L. A. Tsuji
Museum für Naturkunde, Humboldt-Universität zu Berlin, D-10099 Berlin, Germany

A. Turner
School of Biological and Earth Sciences, Liverpool John Moores University, Liverpool L3 3AF, UK

C. G. Turner II
School of Human Evolution and Social Change, Arizona State University,
L. Ulhaas
Forschungsinstitut und Naturmuseum, Abt. Paläoanthropologie und Quartärpaläontologie, Senckenberganlage 25, D-60325 Frankfurt am Main, Germany

P. S. Ungar
Department of Anthropology, University of Arkansas, Old Main 330, Fayetteville, AR 72701, USA

E. S. Vrba
Department of Geology and Geophysics, Yale University, New Haven, CT 06520, USA

G. A. Wagner
Geographisches Institut, Universität Heidelberg, Im Neuenheimer Feld 348, D-69120 Heidelberg, Germany

C. V. Ward
Department of Pathology and Anatomical Sciences, M263 Medical Sciences Building, University of Missouri, One Hospital Drive, Columbia, MO 65212, USA

M. Weingarten
Institut für Philosophie, Philipps-Universität Marburg, Wilhelm-Röpke-Strasse 6, Block B, D-35032 Marburg, Germany

B. Wood
Center for the Advanced Study of Human Paleobiology, Department of Anthropology, The George Washington University, 2110 G St. NW, Washington, DC 20052, USA

E. Zimmermann
Institut für Zoologie, Tierärztliche Hochschule Hannover, Bünteweg 17, D-30559 Hannover, Germany

H. Zischler
Institut für Anthropologie (1050), Fachbereich 10 – Biologie, Johannes Gutenberg-Universität Mainz, D-55099 Mainz, Germany
1 Historical Overview of Paleoanthropological Research

Winfried Henke

Abstract

This chapter provides a comprehensive scientific historical overview on paleoanthropology as a multifaceted biological discipline. A terse compendium on pre-Darwinian theories of evolution is followed by a historical report of the paradigmatic change by Darwin’s perspective on life processes from a teleological to a teleonomic view. Focusing on the fossil discoveries in Europe and later on in Asia and Africa and the different methodological approaches, it becomes obvious that as opposed to other biological disciplines, paleoanthropology remained until post-World War II first and foremost a narrative discipline, apart from the mainstream of biological thinking. Paleoanthropology kept this “iridescent image” in the public opinion until now, wrongly, as is proven. It is shown that since Washburn brought up an innovative conceptual outline on physical anthropology in 1951, there arose a methodological change in the understanding of human evolution as a self-organizing process focusing on the structural and functional adaptations within the order Primates. The anthropological subdiscipline paleoanthropology profited tremendously from this new approach—albeit with some delay. Current paleoanthropological research does not only ask what our forerunners looked like and when, where, and how they evolved but also specifically seeks to explain the processes of radiation, diversification, and variability by concise hypothesis testing. We have to reconstruct the ecological niche of the fossil humans to define the determinants that caused adaptation in human evolution, a process sometimes defined as hominization, including our biogenetic and tradigenetic evolution.

1.1 Introduction

1.1.1 Historical research: more than looking into a mirror!

“Does disciplinary history matter?” This question was asked by Corbey and Roebroeks (2001a p 1) at the eponymous congress focusing on the history of
paleoanthropology and archeology. A negative answer assumes that studying the history of these scientific fields is an unsuitable job, at best a nice leisure time activity for retired colleagues. And since Landau (1991) implies in her book on *Narratives of Human Evolution* that modern paleoanthropologists are still only “storytellers,” a history of paleoanthropology is a double-sided problem: we have to ask whether there is really any need for a dubious historical approach to a biological discipline that only alleges to be doing science.

The allegation that paleoanthropology is not a serious and respectable field of research is not acceptable, although it may have its discreditable aspects (White 2000). In addition, what should one say about the history of science in general? Is disciplinary history no more than a waste of time for practitioners as well as historians of science? Experience shows that there is no uniform answer. However, if paleoanthropologists and historians of science can provide good suggestions for the improvement of the discipline, why should one give up studying the history of sciences or look on this discipline as a small-scale one?

The late Ernst Mayr—an outstanding modern Darwinist—gave a reasonable, albeit not exhaustive, classification of the various historical approaches: (1) elaborated lexicographic histories, (2) chronological histories, (3) biographical histories, (4) cultural and sociological histories, and (5) problematic histories (Mayr 1982). The essential point of all these approaches is that the historicist approach may give rise to reflection as Theunissen (2001) accentuates.

Although some may doubt whether the various approaches suggested (heuristics, source criticism, interpretation, hermeneutics, and analytics) will improve paleoanthropology, Corbey and Roebroeks (2001a) emphasize especially the heuristic value of historical studies. Dennell sees two further lessons from science history; one is awareness of the dangers of fragmentation, the other those of complacency. He considers the lack of dialogue and understanding between the anthropological disciplines as a fatal risk and warns: “There is still excessive specialisation and insufficient dialogue across the disciplines, especially where the terminology and techniques are unfamiliar” (Dennell 2001 p 66). Concerning the aspect of complacency, he explains: “It is always much easier to see the danger in retrospect, but perhaps one of the main lessons to absorb from the history of science is the danger of too many people becoming too comfortable for too long with an idea, just because so many agree with it, and have agreed with it so often in the past” (Dennell 2001 p 66).

These serious and severe arguments for the improvement of science by disciplinary history are valid for every field of study, i.e., “… area of intellectual endeavour that holds a common set of concerns, theories, and procedures or techniques that are intended to address a closely connected web of problems” (Shipman and Storm 2002 p 108).
A major aspect of biological anthropology is that the present is inevitably rooted in the past. For this reason, Dennell (2001 p 65) sees the study of human origins as “... a search for windows that should give us access to the past through fossil specimens, stratigraphic and climatic changes, inferences from the world around us, and the like.” Paleoanthropologists as well as prehistorians and archeologists seek windows to the past but use different sets of keys. A fatal danger is, however, that we may mistake a mirror for a window and simply extrapolate onto the past our own views and prejudices about the present. The important point here is, as L.P. Hartley puts it: “... the past is a foreign country, they do things differently there” (Foley 1987 p 78). For this reason, Foley continues that “The past cannot just be invented or imagined, nor reconstructed solely from observations of the way the world is at the present.” This brings the task and the challenge of paleoanthropological research and other evolutionary sciences to the fore: we are essentially creating models! (Foley 1987; Jones et al. 1992; Henke and Rothe 1994; McHenry 1996; Delson et al. 2000; Henke 2003a, b, 2004, 2005). This kind of approach (Figures 1.1 and 1.2) is by no means narrative but is rather contextual hypothesis testing; and even if in some cases it is scientific speculation, the accent is on the science (White 1988; Henke and Rothe 1994; Henke 2003a). Every scientist needs to arrive at innovative solutions within a channeled imagination.

The gift of Darwin’s theory is “... that there was no blueprint to be followed, only unfoldings of opportunity,” as Howells (1993 p 14) puts it. If the magnificent panoply of life today is the outcome of a real historical-genetic process without a plan, how do we explain the patterns and processes? Darwin’s theory is uncontested as the center of the whole science of biology. In Darwin’s Dangerous Idea, Dennett (1995, blurb) compares the theory of evolution to a universal acid, an imaginary “... liquid that is so corrosive that it will eat through anything!” and asks “Is nothing sacred?” Every traditional concept that tries to explain our existence has to compete with Darwin’s revolutionized worldview. As a consequence of Darwin’s theory of life, even human beings can no longer be explained in a teleological way as expressed by Dobzhansky (1900–1975) in his famous dictum: “Nothing in biology makes sense except in the light of evolution” (Dobzhansky 1973).

Common sense mainly considers paleoanthropology as the study of human fossils and a descriptive and broadly narrative discipline that is dominated by poorly researched and media-friendly “findings” that cause changing views on the process of human evolution. But is the history of paleoanthropological disciplines in fact something more than the demonstration of “... a road full of errors, freak opinions and bizarre concepts finally discarded” (Corbey and Roebroeks 2001a p 1)? Most anthropologists today agree that paleoanthropology is, like other categories of evolutionary biology, a serious subject involving
hypothesis testing and scientific modeling (Foley 1987; White 2000; Wood and Richmond 2000). They are aware that data ascertained from fossils do not speak for themselves and that for this reason the scientific approach involves the creation and testing of hypotheses and theories (Figure 1.3). It is by doubting theoretical insights that we come to questioning; and by questioning, we may perceive the truth—or what we think to be the truth (Popper 1959a, b, 1968, 1983; Foley 1987; Mahner and Bunge 2000; Vollmer 2003). Stronger theories are those that are optimally corroborated; however, even well-tested general theories may conceivably be refuted. Paleoanthropology and other biological subjects are not, like physics, empirical sciences sensu stricto, but this does not mean at all that it is a narrative subject; storytelling is frowned on in every science, and we agree with Kroeber (1953 p 358) who stressed that, “. . . there exists basically only one kind of fundamental science. All genuine science aims at the comprehension of reality, and it uses both theory and evidence, in combination, to achieve this comprehension.”
Like anthropological scientists, paleoanthropologists are only able to establish a reductionistic view of our *conditio humana*, but this is still an essential role as witness of the great scientific and public interest of this discipline. Heberer (1968a) used the terms “Jetzbild” and “Jeweilsbild” (meaning *status quo*) to intimate the changeability of our knowledge and the fact that every paleoanthropological report will be outdated very soon after it is published (Hoßfeld 1997, 2005a, b).
1.1.2 Darwin’s perspective on life processes

The paradigmatic change from a traditional static view into an evolutionary dynamic concept in the beginning of the second half of the penultimate century resulted in the recognition that humans are an integrated part of a historical process. The so-called “Darwinian Revolution,” triggered by Darwin’s masterpiece from 1859, remained in the following decades in a long-lasting state of confusion. As Bowler (1988) demonstrates in his historiography *The Non-Darwinian Revolution*, there were fundamental misunderstandings of Darwin’s ideas. In his judgement, the impact of evolutionism on late nineteenth-century thoughts has been tremendously overestimated by historiographers (Desmond and Moore 1991; Hemleben 1996).

Traditionally, it is assumed that the Darwinian Revolution in biology provided the impetus for a new evaluation of human origins, and Bowler claims that this assumption is valid up to a point. “Because of religious concerns, Darwin and his followers knew that they would have to explain how higher
human faculties had emerged in the course of mankind’s evolution from the apes” (Bowler 1988 p 141). The complex scientific answers within the given frame of ever-increasing biological facts and “disciplinary matrix” sensu Kuhn (Chamberlain and Hartwig 1999) essentially contribute to our self-image and orientation. An argument often given for the scientific necessity of paleoanthropological research is that humans have to know where they come from to decide where to go but “How [do] we know what we think we know” to borrow a phrase from Tattersall’s (1995) eponymous textbook.

Today’s paleoanthropology or human paleontology is a subdiscipline of evolutionary biology that aims to describe, analyze, and interpret the process of human evolution mainly through a vast set of inductive approaches and deductive hypothesis testing (Foley 1987; Henke and Rothe 1994; Wolpoff 1999; Tattersall and Schwartz 2000). If we want to know more about our origins, there are three basic approaches available to reconstruct our evolutionary history (Washburn 1953; Henke and Rothe 1994, 1999a; Wood 2000):

- First, the primatological approach; one can study the closest living relatives to understand the evolutionary context; such work includes field and laboratory work on behavior and cognition and the comparative morphological, physiological, serological, cytogenetical, molecular biological, and genetical studies (Martin 1990; Jones et al. 1992; de Waal 2000; Jobling et al. 2004).
- Second, the paleoanthropological approach; one can reconstruct our evolutionary history from the recovery and analysis of any relevant fossil evidence (Aiello and Dean 1990; Henke and Rothe 1994, 1999a, 2003; Henke 2003a).
- Third, the population genetic approach; one can study the phenetic and genetic intergroup and intragroup variation of recent human populations to provide clues between the geographical samples and their evolutionary history (Jones et al. 1992; Freeman and Herron 1998; Jobling et al. 2004).
- Finally, there is additional information on prehistoric human activities from cultural findings, i.e., the archeological record, and behavioral fossils, e.g., footprints (Klein 1989; Gamble 1999).

Current paleoanthropological research does not only ask what our forerunners looked like and when, where, and how they evolved but also specifically asks, for example, why humans evolved while other primate species died out (White 1988). We have to reconstruct the ecological niche of the fossil humans to define
the determinants that caused adaptation in human evolution, a process sometimes defined as hominization, although some anthropologists argue that this term is misleading (Delisle 2001). When using this perception, we should be aware that it describes a teleonomic process sensu Pittendrigh (1958). In paleoanthropology—as in other life sciences with a chronological perspective—the experiment is the historical process of nature itself. We have correspondingly to interpret this process within the general principles of evolutionary theory ex post factum, and we have to take into account all the problems that arise from the epistemological difficulty known as subject–object identity (Riedl 1975; Mahner and Bunge 2000; Vogel 2000; Vollmer 2003).

1.1.3 Why a scientific historical approach to paleoanthropology?

The present review of the historical development of paleoanthropology as a multifaceted biological subject is intended to focus on its cultural and social background and its many problematic time-specific aspects. This historical point of view promises some explanations concerning the following aspects:

- Which paradigmatic changes in evolutionary thinking guided the field of paleoanthropological research?
- Which cultural, social, and scientific reasons decelerated or accelerated the progress of paleoanthropology?
- Which outstanding scientists and what ideas brought about the integration of biological, geological, and archeological research?
- What was the country-specific impact on the constitution of paleoanthropological research?
- What underlies the successive development from Darwinism to neo-Darwinism, and of the Synthetic Theory and “Modern Synthesis,” respectively to the System Theory of Evolution as the concept and strategy of paleoanthropological research?
- Is paleoanthropology a fossil- and/or media-driven science, triggered by the discovery and public advertisement of exceptional hominin fossils?
- What is the impact of innovative biological techniques in times of multimedia on current paleoanthropology and its image?
- If the accusation is correct that many paleoanthropologists offer poorly researched but media-friendly “findings” (White 2000), how can historic studies on paleoanthropology contribute to a new orientation?
1.2 Paradigmatic change in the nineteenth century: step by step

1.2.1 Darwin’s forerunners, especially in France and England

The biblical view of the permanence of species expressed by Linnaeus’ sentence “Species tot sunt diversae, quot diversas formas ab initio creavit infinitum ens” was the underlying dogma of the Genesis and the Judaic and Christian tradition. Although animal fossils had been described long before Darwin’s theory was published, they were interpreted as witnesses of “lost worlds” within cataclysm models and not as evidence of a real historical-genetic process.

Georges Cuvier (1769–1832), the outstanding French comparative anatomist, did pioneering research on mammalian fossils and contributed to the self-consciously new science of geology. He began to understand that fossils truly represent remains of once-living organisms and argued for the reality of extinction caused by sudden physical events, so-called “catastrophes.” Cuvier first opened up a geohistorical perspective that is now appreciated as his most important legacy to science (Rudwick 1997). Besides this, he adamantly rejected “transformist” explanations. Although he was doubly on the wrong track and his remark “l’homme fossile n’existe pas” slowed the development of thinking on human evolution, Rudwick’s interpretation of the primary texts demystified Cuvier, who was one of the first to professionally plan his research. His approach had an important influence on the scientists of his time.

One of the outstanding opponents of catastrophism was Sir Charles Lyell (1797–1875), British geologist and popularizer of uniformitarianism. He formulated one of the most basic principles of modern geology, the belief that fundamentally the same geological processes that operated in the distant past also operate today. *Principles of Geology*, his specific work in the field of stratigraphy, was the most influential geological work in the middle of the nineteenth century and did much to put geology on a modern footing.

In spite of much progress in natural scientific thinking, the early explanatory approaches of evolutionary theorists were not able to replace traditional views. All pre-Darwinian explanations of diversity and variability were regarded as just another story of natural history because they failed to explain the driving force of evolution.

The evolutionary theory of Jean Baptiste de Lamarck (1744–1829) proved to be a nonvalid explanation for transformation. However, Charles Darwin (1809–1882) looked upon this retrospectively as an “… eminent service of arousing attention to the probability of all changes in the organic, as well as in the
inorganic world, being the result of law, and not of miraculous interposition” (Darwin 1861, preface).

While Lamarck’s theory triggered Darwin’s evolutionary thinking on “transmutation” of species, it was foremost Lyell’s Principles of Geology and Thomas Malthus’ essay on the Principle of Population (which stated that the population size is limited by the food resources available) that inspired Darwin to his multifaceted approach to deciphering the biological principles of evolution. He defined the fundamentals and described evolution as a self-organizing process by a mutation-selection mechanism without the necessity of a creator or deus ex machina. The driving force of this event, natural selection, codiscovered by Alfred Russell Wallace (1823–1913), is the central explanation of the evolutionary process.

Darwin’s and Wallace’s legacy, the theory of natural selection, ultimately led to a paradigmatic change, a totally new view of the development of life systems including human origins. The vintage philosophical questions “Who are we?,” “Where do we come from?,” and “Where are we going?” were transferred from the metaphysical and the philosophical to a biological focus. They created new and important existential questions. Kuhn (1962) saw the truly revolutionary aspect of Darwin’s theory not in its evolutionism but powerful rejection of the traditional, teleological view of nature. How explosive Darwin’s evolutionary theory was, and how conscious Darwin was of this, is reflected in the fact that he hesitated to deal with the questions of human evolution. The first edition of his classical opus On the Origin of Species by Means of Natural Selection, or the Preservation of favoured Races in the Struggle for Life (Darwin 1859) provides proof of this feeling of insecurity. It took him twelve years to publish his ideas in The Descent of Man and Selection in Relation to Sex (Darwin 1871). His innovative focus on human origins set the context for many of the themes of paleoanthropology during the following century. The anthropological challenge within the dynamic evolutionary concept was how to explain ourselves without compromising our posture. Altner (1981a p 3) verbalizes the existential problem as follows: “Der neuzeitliche Mensch ist aus allen ihn übergreifenden Sinnbeziehungen herausgefallen und auf sich selbst und sein Werden zurückgeworfen.”

Evolutionary thinking was widespread during the nineteenth century as various science historians have shown. Bowler (1988 p 5) suggests in this context: “... that Darwin’s theory should be seen not as the central theme in the nineteenth-century evolutionism but as a catalyst that helped to bring about the transition to an evolutionary viewpoint within an essentially non-Darwinian conceptual framework.”

Nowadays we know that most late nineteenth-century evolutionism was non-Darwinian as “... it succeeded in preserving and modernizing the old
teleological view of things”; however, it was a revolution “... in the sense that it required the rejection of certain key aspects of creationism,” as Bowler (1988 p 5) says (see also Desmond and Moore 1991). The heart of Darwin’s materialism, the theory of natural selection, had little impact until the twentieth century. For this reason, the Darwinian Revolution did not take place in the second half of the nineteenth century and remained incomplete until the synthesis with genetics in the twenties and thirties of the last century. But even after this breakthrough, there was no straight scientific approach in paleoanthropology, until anthropologists appreciated the essential corollary of Darwin’s theory that we are only “another unique species” (Foley 1987).

In retrospect, it becomes obvious that the earliest “evidences” of human evolution were disregarded and misinterpreted. Fossils do not speak, but they give silent witness which is a dictum in paleontology. One can only get morphological, ecological, or taxonomical information within a concise methodological approach. Apparently easy paleoanthropological questions about the space and place of human origins do not necessarily have easily obtainable answers, and for this reason it requires rigorous efforts to establish a sophisticated research design and adequate methodology to find solutions for the key questions given earlier (See the section “Why a scientific historical approach to paleoanthropology?”).

As an evolutionary substratum was missing at the time when William Buckland (1784–1856), professor at the University of Oxford, discovered the first human fossil, the “Red Lady of Paviland” at Goat’s Hole in South Wales in 1823, he totally misinterpreted this 26-ka old find (Sommer 2004). Buckland considered the skeleton as of postdiluvian age and was unwilling to attribute any great antiquity to the Upper Paleolithic fossil skeleton.

“For most of the nineteenth century, no one believed or anticipated discoveries that would demonstrate a history of humans and their ancestors stretching back over more than a few thousand years,” write Trinkaus and Shipman (1993 p 9). Although savants struggled with evolutionary ideas for decades, the belief persisted that the human past differed from the present only in the “primitiveness” of the ancient peoples, not in their very essence and being. Starting in the twenties of the nineteenth century, there sprang up one local natural history society after another. The Red Lady inspired the hunt for human fossils and artifacts. Kent’s Cavern in Torquay Devon, today recognized as one of the most important archeological sites in the British Isles, was excavated by John MacEnery and William Pengelly (Sacket 2000). They discovered an Upper Paleolithic skeleton in association with flaked stone tools. John Lubbock, the foremost British archeologist at that time, refused the report on “modern savages” written by Godwin-Austin and considered the findings as “improbable” (Trinkaus and Shipman 1993).
Phillipe-Charles Schmerling, a Belgian physician and anatomist, unearthed an infantile skull in 1829/1830 in Engis near Liège, which was diagnosed later as a Neanderthal fossil. The Engis child was the first specimen of its kind to be discovered (Schmerling 1833). During the same period, Casimir Picard, a physician and avid archeologist, was excavating prehistoric stone tools in France. His passion was experimental approaches to archeology as he attempted to make and use stone artifacts like those that were being found in excavations. Due to inadequate chronology of the artifact-bearing horizons, he was not able to calibrate some of his stone tools as Neolithic. But his experiments led him to conclusions about how the tools were manufactured. Picard developed for the first time systematic excavation techniques and a stratigraphic approach. While his innovative research was without wider impact on archeological progress, he impressed his friend Boucher de Perthes. This most influential aristocrat, who combined his romantic views on human origins with archeological fieldwork, argued for the existence of Pleistocene—or as he said pre-Celtic—humans, but his findings were at first disregarded by the scientific community. When in 1864 some of his findings were published in *The Anthropological Review*, the comment of the scientific board was downright British: “We abstain at the present from offering any comment on the above” (Trinkaus and Shipman 1993 p 44). Although Boucher de Perthes could not avoid the image of flamboyant enthusiast and “madman,” there finally came a change in the assessment of the hand axes as genuine tools. Finally, the precise excavation of the gravel beds of the Somme at St Acheul by the French amateur naturalist Marcel-Jérôme Rigollot (1786–1854) impressed a group of outstanding British colleagues, among others Charles Lyell, and led to the acceptance of the claim that the hand axes were associated with extinct mammal bones—but where did this leave the human fossils? (Klein and Edgar 2002).

1.2.2 Neanderthal case: “Neanderthals without honor”

The Neanderthal man from the Kleine Feldhofer Grotte in the Neander Valley near Düsseldorf was found by limestone workers in 1856 and described by the local teacher Johann Karl Fuhlrott (1803–1877). The fossil was the first early human specimen to be recognized as such. Discoveries like that from Engis (1829) and another from Gibraltar (found in 1848) were made sooner, but their nature became evident much later. Fuhlrott’s merit was that he realized the significance of the fossils, which the limestone workers took for animal bones. Luckily, the owner of the excavation site saved them at the last moment (Schmitz and Thissen 2000). Fuhlrott fought, together with the anatomist
Herrmann Schaaffhausen (1816–1893) who taught at the University of Bonn, for their acceptance as ancient remains from the diluvial age. Both were convinced that the morphological structure of the bones indicated a high-diluvial age. As the discovery of the fossil bones antedated the publication of Darwin’s *Origin of Species*, this specimen has often been apostrophized as first proof for human evolution. A deeper analysis of their contribution demonstrates that both protagonists of paleoanthropological research in Germany were far away from a concise interpretation in an evolutionary sense, although they looked upon their fossils as diluvial forerunner of recent *Homo sapiens* (Zängl-Kumpf 1990).

A contemporary of Fuhlrott and Schaaffhausen was Thomas Henry Huxley (1825–1895), the famous British zoologist, often referred to as “Darwin’s Bulldog” (Desmond 1997). However, Huxley was no Darwinian. Historians know the difficulty of accurate definition, and Hull (1985, 1988) gets to the heart of this problem when suggesting “… that Darwinians are simply those scientists who expressed loyalty to Darwin as the founder of evolutionism, whatever their beliefs about how evolution actually works” (Bowler 1988, p 73). Although Huxley was highly committed by Darwin’s theory, it is well known that he had essential problems with the selection theory. Further, he was strongly tempted by non-Darwinian ideas, e.g., internal factors that would produce changes independent of the environment. Not being really interested in adaptation, Huxley speculated that evolution might sometimes work in a saltatory manner (Bowler 1988; Desmond 1997). His position was that the continuity between humans and other animals does not detract from the inherent specialness of humans. The contemporary so-called proponents of Darwinism were not at all Darwinian evolutionists. In his famous papers on *Evidences as to Man’s Place in Nature*, Huxley (1863) gave morphological arguments for our relationship with recent primates and pointed to the scarce fossil record known in his time. In spite of contrary statements by many historians, Huxley said virtually nothing about human origins but concentrated exclusively on demonstrating the physical resemblances of humans and apes. Concerning the Neanderthal man from Germany, Huxley conducted a sophisticated comparison with anatomically modern skulls from Australian Aborigines and other aboriginal relicts, pioneering new ways of orienting and measuring skulls for easier comparison (Desmond 1997). His conclusion was that the Neanderthal skull emerged as an exaggerated modification of the lowest of the Australian skulls. Huxley stated that the brain of the Neanderthal man was of normal size for an ancient savage, and the stout limbs suggested to him a cold adaptation to glacial Europe. However, in no sense was this specimen “intermediate between Man and Apes.” He viewed the Neanderthals as a very “primitive race” of humans, “the most pithecoid of human crania yet discovered” (Huxley 1863 p 205). The Huxley-biographer Desmond (1997) illustrates this by
reference to Huxley’s diary entries: “Where, then must we look for primeval Man? Was the oldest Homo sapiens pliocene or miocene, or yet more ‘ancient’? How much further back must we go to find the ‘fossilized bones of an ape more anthropoid, or a man more pithekoid’?” Desmond concludes that Huxley was preparing the world for ancient semihumans.

The Irish zoologist William King proposed in 1864 the name Homo neanderthalensis, although his arguments for a separate species in the genus Homo were inadequate, not to say absurd. Since then, opinion has fluctuated as to whether the fossils should be considered as a separate species, H. neanderthalensis or H. sapiens neanderthalensis (a subspecies of H. sapiens). The “fate of the Neanderthals” is the trickiest controversy in paleoanthropology (Henke and Rothe 1999b; Henke 2003a, b, 2005; Finlayson 2004). The question of whether some or all of these fossils deserve a place in our direct ancestry or whether they can be viewed as a single lineage leading to and culminating in the classic Neanderthals of the last ice age is the long-lasting taxonomic problem in paleoanthropology. What became very soon apparent with Huxley’s Evidences was the tremendous need for an extension of the fossil report—since that time colloquially termed as “missing links”—and for an improved comparative methodology to analyze and interpret the fossils of recent primates.

1.3 Theoretical and methodological progress in paleoanthropology since Darwin

1.3.1 Successive discovery of the paleoanthropological background

In Darwin’s time, it was already evident that the scarce evidence of hominin life in the past does not allow the neglect of any clues. We need all available sources to reconstruct our evolutionary history, and these interpretations rely overwhelmingly on fossils. These latter are obviously mute; for this reason, one has to formulate hypotheses about the biological and phylogenetical role of the extinct taxa, and the only reliable approach to increasing our knowledge of the lost worlds is to compare them with recent sets of well-known phenomena.

This scientific process started very soon after the fulguration (sensu Popper) of the pre-Darwinian interpretation of fossil specimens as documents of species of former times and Darwin’s evolutionary theory. The systematic search for phylogenetic forerunners of recent taxa began. The most dramatic part of the Darwinian paradigm focused on the question of how humans evolved from archaic primates. What did the hypothetical species transitional between apes
and man look like? When, where, and how did the missing link live, a term created in 1861 by Asa Gray (Shipman and Storm 2002). Darwin’s evolutionary theory of natural selection did not automatically provide an answer to that question, and Darwin himself was cautious enough not to give a premature answer. But he had provided a new framework in which all these questions could be answered. While the public longed for strong proof of human evolution, Darwin gathered all available arguments for more than 10 years. Finally he published his brilliant anthropological volumes *The Descent of Man and Selection in Relation to Sex* (1871) and *The Expression of Emotions in Animals and Man* (1872). Both books deal with human evolution and particularly with sexual selection, whose enormous evolutionary impact was first understood only around 100 years later, when the sociobiological paradigm emerged (Wilson 1975; Vogel 1982; Miller 1998; Voland 2002; Voland and Grammer 2002).

During the 1860s, Darwin’s ideas were widely popularized. However, there emerged much scientific skepticism at that time as the laws of heredity worked out by Gregor Mendel (1822–1884) in 1865 remained unknown until the twentieth century. Beside Darwin’s British supporters, especially Thomas H. Huxley and Charles Lyell, it were the German geologist and paleontologist Friedrich Rolle (1827–1887) and the zoologist Carl Vogt (1817–1895) who advocated Darwin’s theory (Rolle 1863; Vogt 1863).

Even more committed and sarcastic than the “Affenvogt” was Ernst Haeckel (1834–1919), an outstanding German biologist who sagaciously fought against the “ape complex.” He is best known for his *recapitulation law* (ontogeny recapitulates phylogeny), a highly controversial assumption. After reading *The Origin*, he became a powerful and eloquent supporter of evolution. Although Haeckel admired Darwin’s theory, concerning his own evolutionary thoughts he remained an orthogradualist, and concerning the concept of the “survival of the fittest,” a Lamarckian. Haeckel was not really supportive of natural selection as the basic principle of evolution, and his interest in fossils and paleoanthropology was small. He was convinced that due to the intertwining of phylogeny and ontogeny, ontogenetic structures were sufficient evidence for evolution (Haeckel 1898, 1902, 1905, 1922; Heberer 1965, 1968b, 1981; Hoßfeld and Breidbach 2005; Kleeberg see www).

Despite misunderstanding many of Darwin’s thoughts, Haeckel inspired the public and colleagues with his enthusiasm for evolution and animated the debate. As he was highly motivated by his anti-Christian attitude, his influence in science faded, especially when he created his monistic theories and dabbled in esoteric fields (Hoßfeld 2005b; Kleeberg see www). He gave paleoanthropology an essential impulse by publishing the first phylogenetic tree (Figure 1.4) that
included humankind. Darwin commented on this: “Ihre Kühnheit läßt mich jedoch zuweilen erbeben, aber, wie Huxley bemerkte, irgend jemand muß eben kühn genug sein und einen Anfang machen, indem er Stammbäume entwirft” (Darwin’s letter to Haeckel, November 12, 1868) (Schmitz 1982). As the scarce fossil record only allowed a very hypothetical pedigree, there was much courage needed indeed. Haeckel postulated a forerunner species *Pithecanthropus*
alalus—a speechless ape-man—a missing link, which he believed lived during the Pliocene in Southeast Asia or Africa. Within an orthogenetic pedigree, he posited a primitive species—which he named Homo stupidus—between this “ape-man” and the recent H. sapiens. Around 30 years later, fossils were found in the postulated Asian region by the Dutch physician Eugene Dubois (1858–1940), which roughly fitted the expectation of a Pithecanthropus (Bergner 1965).

Meanwhile a focus of research was the phenomenon of the Ice Ages and the discovery of Upper Paleolithic man and Ice Age cultures (Trinkaus and Shipman 1993; Sackett 2000). Although human fossils were rare at the time of Darwin’s revolutionary discovery, there was much evidence from animal bones, molluscs, sediments, and other materials to give insight into ancient populations. Further, there were ever-increasing indications of long-term fluctuations in the earth’s climate. Geologists, like Agassiz, Geike, and Lyell as well as Brückner and Penck, established a Pleistocene framework of successive glaciations, and archeologists strove to establish the antiquity of human ancestry through the association of stone tools with extinct animals. The geological research resulted in the Alpine model, a chronological system of glaciations that gave a framework for the ongoing discovery of Neanderthal fossils.

As biological, anthropological, geological, and archeological data came together within the frame of the evolutionary theory during the second half of the nineteenth century, many scientific societies were founded, which supported all kinds of scientific research.

Rudolf L.C. Virchow (1821–1902), a famous German physician and anthropologist sensu lato, best known for his guiding research in cellular pathology and comparative pathology, was a universal scientist and liberal politician who founded the German Society of Anthropology, Ethnology, and Prehistory in 1869 (Degen 1968; Andree 1976; Schipperges 1994; Goschler 2002; Trümp 2004). He included humankind in the historicization of nature and came to the conclusion that H. sapiens was “Post-Historie” (Goschler 2002 p 322). For that reason, he was highly skeptical about the validity of Darwinian theory concerning our own species and doubted the phylogenetic classification of the species H. neanderthalensis, as the Irish zoologist William King had dubbed the skeleton from the Kleine Feldhofer Grotte (Stringer and Gamble 1993; Trinkaus and Shipman 1993; Schmitz and Thissen 2000). Due to his “pathologist view,” he interpreted the Neanderthal-derived features or apomorphies as pathological features resulting from arthritis; further from the erroneous information that the skeletal remains were associated with polished stone tools (an indicator for the Neolithic), he concluded that the Neanderthal man must have lived in recent times (Schott 1979). Thus, not all human fossils known at that time were accepted as convincing evidence of our ancestry, especially given their widely uncertain dating.
Neither Virchow nor Haeckel pushed paleoanthropological research, the former from misinterpretation of the facts and skepticism on Darwin, the latter from his conviction that ontogenetic research delivers sufficient information to demonstrate phylogenetic evolution. Virchow’s interest was much more in prehistoric anthropology and ethnology, from the point of view of the decoupling of natural and cultural evolution. The retrograde view that a missing link does not exist was admired by the Christian church, and when *Pithecanthropus erectus*, the so-called “Java Man,” was described, he thought that the bones represented a giant gibbon.

Virchow’s skeptical attitude concerning *Pithecanthropus* as well as his critical interpretation of the Neanderthal fossil as a pathological individual lessened the biological impact of a premature “palaeoanthropology” during the second half of the nineteenth century (Trümper 2004). This holds true for the English and French scientific scenes as well. Although the physician Paul Pierre Broca (1824–1880), the founder of the Société d’Anthropologie de Paris in 1859, was a pioneer of comparative anatomy and anthropology, he never accepted the Neanderthals as fossil documents. His interest was in understanding the patterns of variation in order to understand the significance of anatomical differences. For this reason, he became one of the first to use statistical concepts in establishing anthropology as a scientific discipline in contrast to medical science. In 1882, his French colleagues Quatrefages and Hamy published *Crania ethnica*, a monograph which exemplifies the huge interest in recent cranial variation that dominated the anthropological discipline.

Eduard Lartet (1801–1871) was the first to describe the primate genera *Dryopithecus* and *Pliopithecus*, but much more important was his discovery of signs of prehistoric art made by early humans. The fossil ivory carving of La Madeleine, found in 1864, was presented at the world exhibition in Paris 1867 and raised tremendous interest.

In the following period, neither the discoveries of human fossils from La Naulette in Belgium nor those of Pontnewyydd (Wales), Riveaux (Southern France), Šipka, Mladeč, and Brno (Czech Republic) were able to convince the European scientific community of the human antiquity. Even the analysis of the Spy fossils, which had been discovered in Belgium in 1886, did not slow the rejection of evolutionary thoughts. Trinkaus and Shipman (1993 p 132) summarized: “The man of the Neander Valley remained without honour, even in his own country.”

In the motherlands of evolutionary thinking, France and England, Paleolithic archeology dominated the paleoanthropological discussion and began taking shape as an organized scientific field of research in the 1860s. Sackett (2000 p 38) puts it this way: “… discovering the Paleolithic became a matter of empirically demonstrating that human remains and artefacts could be found in association with the remains of extinct animals belonging to the deep time of earth history.”
Progress was for the first time possible through the scientific recognition of the Pleistocene geological epoch, i.e., the interpretation of fossils within a solid geological background. Although Cuvier and Lyell had done indispensable scientific work to understand the past, the first assumed that the changes came about by a series of revolutions, while the second, although he had major objections to Cuvier’s theory of catastrophism, was no evolutionist either. The challenging question for the evolutionists was how to provide empirical data concerning the existence of “diluvial” and “antediluvian” early human populations. As these workers had to rely on the research of the paleontologists and the archeologists, paleoanthropological science was becoming multidisciplinary even at this early stage. The evidence from the Paleolithic record was mainly from data from stream gravel terraces, rock shelters, and bone caves. As Sackett (2000 p 42) mentions, from the 1820s until 1859 there was a series of discoveries in France (especially in the Perigord), England (Paviland bone cave near Swansea, Wales; Kent’s Cavern, near Torquay, Devonshire), and Belgium (Engis near Liège). The research resulted in the assignment of Pleistocene fauna to deep geological times, but it did not support evolutionary thinking due to the many alternative explanations available. The consequent question was whether actual human remains existed in association with Pleistocene animal fossils and undoubted artifacts. The immediate question of human antiquity came up with the discussion of the fossils from the Neanderthal in Germany and Darwin’s evolutionary theory. Most of the geologists and paleontologists remained skeptical; they saw no proof of a high antiquity of humankind from evidence from the bone caves or gravel terraces. One reason for their skepticism resulted from the archeological work of Jacques Boucher de Perthes, director of customs at Abbeville, France, whose empirical evidence was much doubted. As consequence of his having “... found much too much,” as Sackett (2000 p 45) puts it, the Bible continued to dominate everyday metaphor; and ironically enough, the successful archeological research in the Near East, Egypt, and Palestine solidified the traditional view that human-kind was unique and doubtless recent. In spite of many convincing facts—as retrospectively gauged—from archeology, the social establishment of the mid-nineteenth century maintained the older view.

An opportunity for change came in 1858, when Brixham Cave near Torquay on the Devon Coast was discovered, and outstanding scientists like Hugh Falconer (1808–1865), Charles Lyell, Richard Owen (1804–1892) supervised excavations there. William Pengelly (1812–1894), a local schoolteacher and geologist, was able to gather, by a new method of layer-by-layer excavation, thousands of animal bones including those of hyena, cave bear, rhinoceros, and reindeer. The impact of these fossils on the question of human antiquity would have been zero if Pengelly had not found undeniable artifacts, which he
described as “knives”. These chipped stones from Brixham Cave challenged the received opinion about human antiquity in 1859. While the belief in human antiquity of the excavator and some colleagues was confirmed by the association of fossil bones and the hand-axes, others like Owen did not agree and thought that the animals had not become extinct until geologically modern times.

New aspects came into the stagnating discussion when the implements of the Brixham Cave were compared to the finds at Abbeville and Amiens, which had been excavated by Boucher de Perthes. The ultimate convincing facts came when the English team, digging in the Somme terraces, was able to document a hand-axe in place in a fossil-bearing stratum at St Acheul. This was the turning point as Pengelly, Prestwich, and Lyell were then able to convince the British establishment of the antiquity of humankind. The French scientific community, which had contradicted Boucher de Perthes’ interpretations for many years, now no longer rejected the idea of human antiquity. However, some French scientists, like the zoologist Isodore Geoffrey St Hilaire (1805–1861) or Eduard Lartet (1801–1871), had been more or less convinced about human antiquity before, but the discovery of human teeth intermixed with fossils of cave bear and hyena in a cave near Massat in southern France brought about final acceptance. Lartet published the ultimate proof in 1860. He described cut-marks on fossil bones that had been made by stone tools when the bones were still fresh. This was the essential evidence of contemporaneity of humans and extinct animals, and it signaled the start of intensified geological research. The glaciological research aimed at structuring the Pleistocene epoch helped to put the human antiquity into a chronological frame; and the amalgamation of archeological and geological facts brought up a new era that started with Lyell’s first edition of his famous *Geological Evidences for the Antiquity of Man* (1863). The coming together of diverse aspects of the cultural and natural sciences ultimately yielded an innovative conception of man and his origin (Daniel 1959, 1975; Trinkaus and Shipman 1993; Sackett 2000; Murray 2001).

### 1.3.2 Southeast Asia as supposed cradle of humankind

If paleoanthropologists had to answer the question of which fossils had the most exceptional influence on human evolutionary thinking, they would, of course, include the *Homo erectus* fossils from Java alongside those from the Neander Valley. The assessment is very easy to understand: first, there is a hero, a young, enthusiastic physician from the Netherlands, Eugene Dubois, who feels inspired by Darwin’s theory and Haeckel’s preliminary draft of a link between the lesser apes and earliest human populations; he joins the army as military surgeon and embarks for Indonesia. Beside his service in the army, he is looking for the so-called
missing link, *P. alalus*, as Haeckel had dubbed the speechless and small-brained human species (Shipman 2001; Shipman and Storm 2002). Second, this paleontological amateur is successful in discovering a fossil human tooth and then a skullcap in the gravels of the Solo River. He proposes at first the name *Anthropopithecus* and then changes it to *Pithecanthropus*, the name given by Haeckel. Finally, a stroke of good fortune leads to the discovery of a human femur, which seems to belong to the *Pithecanthropus* fossils, which causes the species to be named *erectus*. Third, Dubois, the man who found the posited missing link, gets into more and more trouble with his critics and, due to severe personal problems, the end of his career is tragic. Since Dubois was neither an accidental discoverer nor a paranoid eccentric person, as science historians and colleagues (v. Koenigswald 1971) have sometimes characterized him, so the reinvestigation of his lifework gives a less discrediting view (Shipman 2001; Shipman and Storm 2002).

If one takes all aspects of this story together, it shows the main characteristics of a myth. On the one hand his unbelievable luck and on the other hand, his unyielding, difficult character and his stubborn nature made Dubois an interesting public figure. While some historians like Erickson (1976) look upon Dubois as an unimportant figure in paleoanthropology, others like Haddon (1910), Theunissen (1981), or Shipman (2001) see him as one of the founding fathers of paleoanthropology (Shipman and Storm 2002).

Did Dubois’ essential discovery coincide with a paradigm shift in paleoanthropology? While Howell (1996 p 4) stated that the entire field of paleoanthropology is “... close to a paradigm state without yet having achieved it,” Chamberlain and Hartwig (1999 p 42) suggest, “... that the validity and promotion of knowledge claims in palaeoanthropology are explicitly Kuhnian and that future epistemological progress depends on acceptance of both Kuhnian and positivist approaches to knowledge-building.”

In opposition to Chamberlain and Hartwig’s opinion, Cartmill (1999 p 46) stated that: “Whatever a paradigm is, normal science consists of attempts to ‘articulate’ a paradigm by answering questions, verifying predictions, and solving sticky problems in paradigmatic terms.” He continues: “Conceptual change in science is perpetually going on at all levels, and is not ordinarily concentrated in punctuational events that can be distinguished as ‘revolutions’.”

Whatever position one adopts, there are obviously good reasons to regard Dubois as a founder of paleoanthropology. According to Shipman and Storm (2002 p 109), founders must have special qualifications, e.g.,

- A pivotal discovery of fossils or artifacts
- Development of an innovative technique for the description and analysis of discoveries
A new framing of the problems or questions of the fledgling field
A broad dissemination of information or debate about the subject, which serves to make it a matter of wide concern
The provocation of a general reaction or response from potential colleagues.

The last point fits with the Machiavellian fact that an invention needs the attention of the scientific community to become common currency. One may ask what would have happened to Dubois’s discovery in modern times. The rules of “The Economy of Attention,” formulated by Georg Franck (1998a, b), list the following stepwise pattern: “Attention by other people is the most irresistible of drugs. To receive it outshines receiving any other kind of income. This is why glory surpasses power and why wealth is overshadowed by prominence.”

While Haeckel, Broca, and other scientists focused on the comparative anatomy and embryology as proof of human evolution, Dubois was the first who implemented a new strategy. Since he realized that human evolution is a chronological process, which must have happened in preferred localities with specific ecological niches, he set out for Southeast Asia, promised by Haeckel’s fictive pedigrees as a possible “cradle of mankind” where the mysterious missing link should have lived.

Dubois was the first to write a detailed monograph on a hominin fossil, and he applied for the first time metrical and mathematical procedures for the calculation of brain volumes and stature heights to a human fossil. A critical evaluation of his work by Shipman (2001) shows that his scientific work was innovative and erudite, contrary to reports circulated by paleoanthropologists and science historians. The contemporary reviews of his 1895 monograph by, e.g., the German anatomist Wilhelm Krause and the multitalented scientist Rudolf Virchow were scathing, just as were those of the British, French, and Swiss scientists. On the other hand, it was no surprise that Haeckel’s comment was positive. The skepticism of part of the scientific community incited him to tremendous activity, and there resulted a flow of papers on *Pithecanthropus*. It is thanks to Dubois that the focus on human fossils increased (Shipman and Storm 2002). The market place for human fossils and paleoanthropological discussions was open (Sander 1976).

Among the scientific activities of the anthropological organizations in different European countries, e.g., the Gesellschaft für Anthropologie, Ethnologie und Urgeschichte in Berlin or the Société d’Anthropologie de Paris as well as the Royal Anthropological Society of Great Britain, at the turn of the penultimate century, interest in the diversity of recent humankind was paramount. While many activities were concentrated on the typological classification of recent populations as well as prehistorical and archeological research, paleoanthropological
problems and fossils played only a minor role. Dennell (2001 p 52) designates the period from ca. 1870 to 1930 as the “Age of Prejudice so far as European (and North American) perceptions of non-whites is concerned.” The dominating theme was the racial prejudice of fixity and inequality of races. While this attitude did not change when the first non-European fossil caught the interest of the scientists and the public in general, it introduced a new aspect of human evolution: the confrontation of eurocentristic typological views with paleoanthropological facts.

In the light of the *Pithecanthropus* fossils, the Neanderthal problem reached a new dimension. Virchow’s diagnosis implied that the Neanderthal man was diseased with rickets as a child and arthritis as an adult. To many, this explanation of the special bony features of Neanderthals sounded far-fetched (Trinkaus and Shipman 1993). There was a fundamental need for appropriate comparative biological research and a reasonable taxonomic approach. With the benefit of the hindsight, one can see that in those times there was no theoretical basis for elaborated paleoanthropological research, since the anatomists as well as the archeologists were concentrating on case studies, while zoologists, like Haeckel, were focusing on other topics than merely paleoanthropology. The end of the nineteenth and the beginning of twentieth century yielded exciting discoveries of human fossils (*Table 1.1*). The obvious lack of a sophisticated theoretical basis

<table>
<thead>
<tr>
<th>Year</th>
<th>Place and site</th>
<th>Specimens, taxon, pathbreaking finding</th>
<th>Aspects and comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1830</td>
<td>Engis, Belgium</td>
<td>First Neanderthal fossil, Mousterian culture</td>
<td>Described much later as such</td>
</tr>
<tr>
<td>1848</td>
<td>Forbes’ Quarry, Gibraltar</td>
<td>Neanderthal calvarium</td>
<td>Described much later as such</td>
</tr>
<tr>
<td>1856</td>
<td>Kleine Feldhofer Grotte (Neanderthal, Germany)</td>
<td>Calotte and postcranial skeleton, no archeological remains</td>
<td>First human fossil remains which have been attributed by Fuhlrott and Schaffhausen in 1857 as diluvial remains, detailed description by Fuhlrott in 1859</td>
</tr>
<tr>
<td>1859</td>
<td>London (UK)</td>
<td>Charles Darwin publishes his “Origin of Species by Means of Natural Selection …”</td>
<td>Darwin’s descent theory and his explanatory theory of selection induced a paradigmatic change</td>
</tr>
<tr>
<td>1866</td>
<td>La Naulette (Belgium)</td>
<td>Mandible, Würm I</td>
<td>Further indication for the existence of fossil man</td>
</tr>
</tbody>
</table>
### Table 1.1 (continued)

<table>
<thead>
<tr>
<th>Year</th>
<th>Place and site</th>
<th>Specimens, taxon, pathbreaking finding</th>
<th>Aspects and comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1868</td>
<td>Cro-Magnon (France)</td>
<td>Fossil remains of several individuals, best known as “Le Villard” (Cro-Magnon 1), type specimen of a so-called “Cro-Magnon race” within contemporary typological classifications of those times</td>
<td>The fossils have been described by Vacher de Lapouge as <em>H. spelaeus</em> but are fully modern Upper Paleolithic humans. The association of the “Cro-Magnon”-Type with an (evolved) Augrinacien and an extinct Pleistocene fauna was essential for the acceptance of a human antiquity</td>
</tr>
<tr>
<td>1880</td>
<td>Šipka (Czech Republic)</td>
<td>Fragmentary mandible of a child, Mousterian, Würm I/II</td>
<td>The Neanderthal mandible did not convince the critics of the evolutionary theory</td>
</tr>
<tr>
<td>1886</td>
<td>Spy (near Namur, Belgium)</td>
<td>Two skulls and postcranial remains</td>
<td>The morphology of the bones disproved any doubt on the existence of Neanderthal man in Europe</td>
</tr>
<tr>
<td>1887–1892</td>
<td>Taubach (Germany)</td>
<td>Isolated teeth, Mousterian culture</td>
<td>“Pre”-Neanderthal remains</td>
</tr>
<tr>
<td>1891–1898</td>
<td>Trinil (Central Java, Indonesia)</td>
<td>Calotte and complete femur, partial femur, teeth</td>
<td>The discovery of the Trinil fossils by Eugène Dubois (<em>Pithecanthropus erectus</em>) was essential for the discussion of the “missing link” and displaced the focus on human origins from Europe to Asia</td>
</tr>
<tr>
<td>1899–1905</td>
<td>Krapina (Croatia)</td>
<td>&gt;670 cranial remains, Riss-Würm-Interglacial, cannibalism, burnt bones, Mousterian culture</td>
<td>The “<em>H. primigenius</em>” (later on <em>H. neanderthalensis</em>) was excellently described and analysed by Gorjanović-Kramberger (1907)</td>
</tr>
<tr>
<td>1907</td>
<td>Mauer (near Heidelberg, Germany)</td>
<td>Fossil mandible, dated to the middle Pleistocene, between the Cromerian and Holsteinian interglacials, ca. 500 ka</td>
<td>Schoetensack (1908) described the Mauer jaw as a new species, <em>H. heidelbergensis</em>. Some regard the fossil as <em>H. erectus</em> or classify the specimen as archaic <em>H. sapiens</em></td>
</tr>
<tr>
<td>Year</td>
<td>Place and site</td>
<td>Specimens, taxon, pathbreaking finding</td>
<td>Aspects and comments</td>
</tr>
<tr>
<td>------------</td>
<td>---------------------------</td>
<td>--------------------------------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>1908</td>
<td>La Chapelle-aux-Saints</td>
<td>Well-preserved Neanderthal skeleton with a lot of pathologies, Charentian Mousterian, Würm II</td>
<td>Boule’s description of the fossil (1911–1913) was a milestone of research of the Neanderthals and “established” them as a separate species of their own; long time regarded as “archetype” of the classic Neanderthals of western Europe</td>
</tr>
<tr>
<td>1908</td>
<td>La Quina (France)</td>
<td>Remains of a total of 27 hominine individuals, highly fragmentary, including an infant skull (H 18); Quina variant of the Mousterian, traces of fire</td>
<td>Louis Henry-Martin described the fossil material in 1908 and in the following years as classic Würm Neanderthal remains</td>
</tr>
<tr>
<td>1908/1914</td>
<td>Le Moustier (France)</td>
<td>Adolescent and infantile skeleton, associated with Mousterian culture and overlying Châtelperronian assemblage</td>
<td>Described by Klaatsch and Hauser (1910) as the type specimen of the species H. mousteriensis, although the Neanderthal affinity of the specimens has never been disputed</td>
</tr>
<tr>
<td>1908–1915</td>
<td>Piltdown (Sussex, England)</td>
<td>From “Pleistocene gravels” “unearthed” specimen, a chimera of an intentionally manipulated mandible of a juvenile orangutan and portions of the skull of an anatomically modern man.</td>
<td>The specimen was named Eoanthropus dawsoni and caused a lot of trouble in paleoanthropology as most of the leading British paleoanthropologists regarded the fossils authentic</td>
</tr>
<tr>
<td>1908–1913 and 1914/16/25</td>
<td>Weimar-Ehringsdorf (Germany)</td>
<td>Calvaria, parietal, mandible, teeth, older than the classical Neanderthal man, Mousterian culture</td>
<td>“Pre”-Neanderthal remains</td>
</tr>
<tr>
<td>1909–1921</td>
<td>La Ferrassie (France)</td>
<td>Fossil material of two adult and several highly incomplete and poorly preserved immature individuals, crouched burials</td>
<td>Field descriptions by Capitan and Peyrony (1909, 1911–1912) and detailed analysis by Boule (1911–1913)</td>
</tr>
</tbody>
</table>
### Table 1.1 (continued)

<table>
<thead>
<tr>
<th>Year</th>
<th>Place and site</th>
<th>Specimens, taxon, pathbreaking finding</th>
<th>Aspects and comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1918 f.</td>
<td>Zhoukoudian (near Beijing, China)</td>
<td>Zdansky collected in the large karst cave human teeth and Bohlin started the excavation campaign 1928–1929 followed by further excavations</td>
<td>The successful excavation of the site shifted the focus of paleoanthropological research and the origin of mankind entirely to Asia. Black’s description of <em>Sinanthropus pekinensis</em> established a species very similar to <em>P. erectus</em></td>
</tr>
<tr>
<td>1921</td>
<td>Broken Hill (newly Kabwe; near Lusaka) Zambia (formerly Rhodesia)</td>
<td>The well-preserved calvarium is associated with an African MSA, dated to 200–125 ka</td>
<td>Woodward (1921) classified the “Rhodesia Man” as a new species <em>H. rhodesiensis</em> and Pycraft (1928) opted for a new genus <em>Cyphanthropus rhodesiensis</em>, while Mourant (1928) saw Neanderthal affinities. Actually the specimen is mostly attributed to <em>H. heidelbergensis</em></td>
</tr>
<tr>
<td>1924</td>
<td>Taung (near Bophythatawana, South Africa)</td>
<td>The skull and hemiendocast of a child have been described by Raymond Dart (1925) in a <em>Nature</em> article as <em>Australopithecus africanus</em></td>
<td>The characters in Dart’s <em>A. africanus</em> were diametrically opposed to those expected; for this reason his classification was highly criticized by Woodward and Keith, the contemporary leading British paleoanthropologists</td>
</tr>
<tr>
<td>1924–1926</td>
<td>Kiik-Koba (Crimea, Ukraine)</td>
<td>Hominid remains including an adult male and an infant (possible burials) assigned to <em>H. neanderthalensis</em>, Mousterian occupation level, Würm I</td>
<td>Extension of the distribution map of Neanderthal sites to eastern Europe</td>
</tr>
<tr>
<td>1926</td>
<td>Gánovce, Slovakia</td>
<td>Cranial and postcranial remains, Riss-Würm Interglacial, Mousterian culture</td>
<td>“Pre”-Neanderthal remains</td>
</tr>
<tr>
<td>1927</td>
<td>Mugharet el-Zuttiyeh, near Lake Galilee, Levante (today Israel)</td>
<td>Frontal skull fragment with an intermediate morphology, neither modern nor Neanderthal</td>
<td>First human fossil remain from the crossroads of the Middle East Corridor</td>
</tr>
<tr>
<td>1929</td>
<td>Saccopastore near Roma (Italy)</td>
<td>Clavarium (Saccopastore 1), skull fragments, upper jaw, Riss-Würm-Interglacial</td>
<td>“Pre”-Neanderthal remains from stage 5e (130–120 ka)</td>
</tr>
</tbody>
</table>

and elaborate methodological skills in paleoanthropological research gradually led to the development of a new field of evolutionary biology.

The best basis for such a development existed in France, where Pierre Marcelin Boule (1861–1942), who was a qualified geologist, paleontologist, and archeologist, unified all necessary attributes to establish a paleoanthropology. His classical description of the Neandertal skeletons from La Chapelle-aux-Saints (Boule 1911–1913) was a landmark in the history of human paleontology (Heberer 1955a). Contemporaneously, the German anatomist Gustav Schwalbe (1844–1916) analyzed a skull fragment from Eegisheim as well as the famous Javanese *Pithecanthropus erectus* (Schwalbe 1906; Fischer 1917): research that is regarded as critical to the founding of paleoanthropology as a research field of its own. Heberer (1955b p 298) comments on this: “An die Stelle ungenügend fundierten Theoretisierens tritt jetzt in der menschlichen Fossilforschung die exakte Empirie.” Albeit this sounds somewhat effusive, the empirical approach to the morphological and metrical analysis of the fossils allowed an independent development, separated from geology and archeology. This view characterizes, on the one hand, the lacking cooperation of paleoanthropology as a biological field of research with the geosciences as well as the archeology/prehistory and, on the other hand, the overestimation of the scientific importance of quantifying morphological methods (Chaoui 2004; Hoßfeld 2005b). Both of these aspects hampered any essential progress in paleoanthropology for decades, the subject remained mainly narrative and descriptive (see Henke, 2006).

### 1.3.3 Development of principles and methodical skills in paleoanthropology

Neither Haeckel nor Virchow regarded paleoanthropology as a particularly important subject, and even anthropology itself was not a discipline of great significance in their times. The reason for the late establishment of anthropology as a separate biological discipline is, as Grimm (1961 p 1) suggested, that “... der Anthropologe viel weniger als der Zoologe oder der Prähistoriker oder der Anatom in der Lage schien, die Grenzen seines Faches zu bestimmen.”

In the first half of the twentieth century, paleoanthropology remained uncentered as physicians, anatomists, biologists, archeologists, and ethnographers from different viewpoints and with different intentions and aims all practiced anthropology. This becomes especially obvious if one looks from the more zoological fields of research across to archeology. The division of so-called explanatory natural sciences (erklärende Naturwissenschaften) from the comprehending humanities (verstehende Geisteswissenschaften) *sensu* Dilthey 1883...
(Groethuysen 1990) was completed at the beginning of the twentieth century, contingent on the obvious consciousness of superiority of natural sciences (Figure 1.5).

Figure 1.5
Paleoanthropology versus archeology—identical aims but different approaches (Henke and Rothe 2006)

Paleoanthropology

Archeology

Identical aims of research:

Reconstruction of paleoecological scenarios, i.e., of environments, social structures, and ecological circumstances of fossil species and populations

Biological substrate —
fossil material
biogenetical research material

Cultural remains

Natural sciences
“erklärende Wissenschaft”
“explaining science”

Cultural / Social sciences
“verstehende Wissenschaft”
“comprehending science”

The rediscovery of Gregor Mendel’s laws in 1900 by Carl Correns, Erich von Tschermak-Seysenegg, and Hugo de Vries gave a push to evolutionary thinking. Around 1930, protagonists of neo-Darwinism like Ronald A. Fisher, John B.S. Haldane, and Sewall Wright developed around 1930 the basic principles of population genetics in the USA and induced a fruitful diversification of natural sciences. In the following period, the Russian geneticist and evolutionary biologist Theodosius Dobzhansky (1900–1975), the American paleontologist George Gaylord Simpson (1902–1984), the German evolutionary taxonomist Ernst Mayr (1904–2005), and finally the British biologist Julian Huxley (1887–1975) founded the synthetic theory of evolution, integrating additionally quite a number of neighboring disciplines to reconstruct the phylogenetic process of our own origin (Jahn et al. 1982; Jahn 2000; Junker 2004; Wuketits and Ayala 2005; Hoßfeld 2005b). The major principles of the Synthesis itself were subsequently enunciated
between 1937 and 1944 in three seminal books. While Dobzhansky (1937) focused on the gene, Mayr’s (1942) vantage point was the species and Simpson’s (1944) perspective was on the higher taxa. Tattersall’s (2000a p 2) assertion that “. . . the Synthesis was doomed to harden, much like a religion, into dogma whose heavy hand continues to oppress the science of human origins a half-century later” is controversial as is reflected in Foley’s statement (Foley 2001). However, before we discuss the consequences of the Synthesis, we shall have a closer look at paleoanthropology.

### 1.3.4 Fossils: hypotheses, controversies, and approaches

In the early beginning of paleoanthropology the main question was quite simple: is there a fossil record which proves the existence of our ancestors from ancient times? The protagonists of paleoanthropology soon recognized the need for a sophisticated empirical approach. The best basis for such a development existed in France, where Pierre Marcelin Boule (1861–1942), a qualified geologist, paleontologist, and archeologist, unified in persona all necessary attributes to establish a paleoanthropology. His classical processing of the Neandertal skeletons from La Chapelle-aux-Saints (Boule 1911–1913) became a landmark in the history of human paleontology (Heberer 1955a). He aimed to understand the patterns of variation and the significance of anatomical differences. For this reason, Boule invented special instruments for quantification(s) and simple statistical concepts to analyze the variation in human skeletons (Boule 1921, 1923; Boule and Vallois 1946, 1952). Boule established a paleontology of humans, later on called paleoanthropology, as a scientific discipline, but as evolution in itself was still regarded as a widely speculative myth, the debate on human evolution “… rose and fell like a tide in France, Germany, England, and the United States,” as Trinkaus and Shipman (1993 p 154) described the situation.

During the late nineteenth and the early twentieth century, the overestimation of quantifying procedures strongly affected the anthropological research, starting with the first representatives of anthropological chairs, Jena Louis Armand de Quatrefages de Breau (1855, Paris) and Marcelin Boule (1867, Paris), and Johannes Ranke (1886, Munich) and Felix v. Luschan (1900, Berlin). While Ranke was deeply opposed to the idea of human evolution, with Otto Schwalbe, just mentioned, Theodor Mollison (München), Felix v. Luschan (Berlin), Rudolf Martin (München), and Otto Schlaginhaufen (Zürich) successively advocated measuring techniques of every conceivable kind (Keller 1995; Chaoui 2004; Junker 2004; Hoßfeld 2005b). While Schwalbe was convinced that information on evolutionary history could only be gained from neo- and paleozoology, others
focused predominantly on typological classification of prehistorical populations and neglected broader paleoanthropological questions. This can only be understood by the fact that the conventional wisdom of that time was that modern “races” arose from types that were already separate in the early Pleistocene. For this reason, there was a big problem when Gorjanović-Kramberger (1906) described the fossils from Krapina, excavated in 1899–1905, in a voluminous monograph as belonging to a “man-eating” population (Radovčić et al. 1988). As this thesis was accepted by one of the most outstanding anthropologists at those times, Aleč Hřdlicka (Prague, later on Washington) who pushed paleoanthropological research using his influential position at the Smithsonian Institution, Krapina became a big challenge to the scientific community (Henke 2006).

Further controversies were born when descriptions of the Upper Paleolithic skeletons from, e.g., Chancelade, Grimaldi, Combe Capelle, Oberkassel, and the Neanderthals, as well as, e.g., Le Moustier, La Chapelle-aux-Saints, La Ferrassie, La Quina were excavated and published. There was special interest in the mandible from Mauer that was described in 1908 by Otto Schoetensack (1850–1912) as Homo heidelbergensis (Adam 1997; Wagner and Beinhauer 1997).

In England, the motherland of Darwinism, there was for a longtime little progress in paleoanthropology as the Darwinists were obviously “no Darwinians in the modern sense” (Bowler 2001 p 14). During the period from 1860 to 1940, a Darwinian style of explanation began to replace non-Darwinian developmental models stepwise, but a real breakthrough or turnover did not occur before World War II, largely due to the overwhelming influence of the paleontologist and geologist Arthur Smith Woodward and the zoologist Sir Arthur Keith as the leading authorities on human remains. Keith published An Introduction to the Study of Anthropoid Apes in 1896, followed by a monograph, The Antiquity of Man, which appeared in 1915. This publication on all-important fossil human remains founded his worldwide reputation and appeared in an enlarged edition in 1925. As can be seen from his textbook Concerning Man’s Origin (Keith 1927), he was at that time less convinced than he had been 10 years before that modern humans and extinct “primitive” types had lived contemporarily. In his volume New Discoveries, Keith (1931) rejected this interpretation given the strength of the arguments for an evolutionary branching of hominins. From the viewpoint of history of science, we have to ask why the leading “anthropologist” as well as his compatriot Sir Grafton Elliot-Smith (1871–1937) retained this view for such a long period, in the face of compelling arguments to the contrary from paleoanthropological discoveries in Europe, Asia, and after 1925 in South Africa too.

One plausible explanation is that theories are the filters through which facts are interpreted as Popper (1959b) said. The reason why the English authorities...
adhered to wrong models combines the fatal misinterpretation of the Piltdown hoax with wishful thinking and cultural bias (Spencer 1990a).

The background of the Piltdown man forgery has been analyzed and described in many ways (Spencer 1990a, b). The most plausible explanation why the *Eoanthropus dawsoni* hoax was so successful is that it seemed to provide proof for a missing link between apes and humans, using a mix of plesio- and apomorph characters. Especially, the primitive jaw and dentition made *Eoanthropus* a more suitable intermediate candidate than *Pithecanthropus*. Even after the discovery of *Australopithecus* at Taung in 1924 and the excavation of the first *Sinanthropus* skull from Zhoukoudian (Chou-kou-tien) shortly thereafter, Keith (1931) hypothesized that the Piltdown type arose from the main ancestral stem of modern humanity.

At that time, prominent British anthropologists, such as Smith Woodward, Keith, and Elliot Smith, were fixed on a European origin of humankind and were absolutely in opposition to models of Asian and African origin. The expected phylogenetic sequence was that the cerebralization antedated the changes of the viscerocranium. As the plesiomorph jaw and the apomorph brain of *Eoanthropus* complied with this expectation, it fitted perfectly into a scheme that was in fact wrong.

It took more than 40 years until J.S. Weiner, Sir Kenneth Oakley, and Sir Wilfrid Le Gros Clark jointly exposed the hoax, although there was much skepticism and rumor earlier. Whoever the players were in this black mark in science, they were aware of the attractiveness and fascination of fossils, the rare resources that help to decipher our place in nature, and they obviously knew about the public appeal (Stringer and Gamble 1993; Walsh 1996; Weiner and Stringer 2003; overview in Harter www; Turritin www). The interpretation of the Piltdown fossil as a human precursor was partially responsible for the vehement dismissal of the first *Australopithecus* from South Africa. Dart’s interpretation of the Taung child as missing link between ape and man yielded a storm of controversy (Tobias 1984). As Trinkaus and Shipman (1993 p 206) put it: “The entire scientific coterie of Britain believed in the fossils without hesitation.” This is remarkable insofar as Dart’s discovery matched the prophecy of Darwin (1871 p 202): “It is, therefore, probable that Africa was formerly inhabited by extinct apes closely allied to the gorilla and chimpanzee; and, as these two species are now man’s nearest allies, it is somewhat more probable that our progenitors lived on the African continent than elsewhere.”

To the extent that these and other indications for an extra-European “cradle of humankind” were deliberately ignored, the Piltdown case is a telling example of cut and dried opinions. Small wonder that the scientific output of paleoanthropology during the first decades of the last century was very heterogeneous and confusing.
and more redolent of stagnation than progress. The evolutionary biology at those times was characterized by Mayr as “chaotic” (Tattersall 2000a p 2).

First, even at the beginning of the twentieth century, Darwin’s principles were widely misunderstood by anthropologists, who persisted in orthogenetic biological thinking or insisted on the theoretical split between natural sciences and humanities. Paleoanthropological theory and methodology were still in statu nascendi, judging from the literature of the time (Klaatsch 1899; Hrdlička 1914; Werth 1921; Weinert 1928; Wiegers 1928; Abel 1931; Hooton 1931; Keith 1931; Le Gros Clark 1934).

Second, the rediscovery of Mendel’s work in 1900 demonstrated that new hereditary variation, i.e., mutation, occurs in every generation and every trait of an organism. The founding of genetics at the beginning of the twentieth century gave tremendous support to Darwin’s theory and removed the objections to his postulates. However, the effects of genetics and population genetics were felt less in paleoanthropology than in the biology of recent populations, and especially in aberrations like typological social biology, ethnogeny, and race typology. In Germany, social Darwinism and fatal concepts of hereditary and race dominated biological and medical research on living people. Paleoanthropology was of only minor interest, while the aim of anthropology was defined in the National Socialist area or “The Third Reich” “... der deutschen Volksgemeinschaft zu dienen” (Reche 1937; see further Zmarzlik 1969; Gould 1983; Müller-Hill 1988; Seidler and Rett 1988; Weingart et al. 1988; Zängl-Kumpf 1992; Hoßfeld and Junker 2003; Junker 2004; Hoßfeld 2005a, b).

Third, various approaches to explain the process of hominization later on turned out to be politically highly incorrect: e.g., Kollmann created in 1885 the term “neoteny” suggesting that humans evolved from pygmies who had simply retained juvenile features during size increase (Kollmann 1902). The basic idea was that pygmy progenitors probably arose from juvenile apes that had lost the ancestral tendency to regress. Bolk (1926) argued in a long series of papers that man evolved by retaining the juvenile features of his ancestors. His “fetalisation hypothesis” influenced evolutionary thinking on human origins and formed expectations of the appearance of transitional species. Bolk’s theory was highly criticized and later on rejected by Starck (1962) because it was built on flawed arguments and a misinterpretation of Darwin’s principles. Actually similar questions still resound in the current so-called evo-devo-discussion, e.g., the evidence overwhelmingly suggests that neoteny, the retention of juvenile characteristics, was one of the most important processes involved in the origin of *H. sapiens* (Mingh-Purvis and McNamara 2002; Chapter 3 of this volume).

Fourth, the discovery of numerous Neanderthal skeletons (Table 1.1) caused new arguments in the discussion of gradualism versus continuity. Boule’s monograph evicted the Neanderthals from our family tree, although some outstanding
early twentieth-century paleoanthropologists, like Aleč Hrdlička, argued for gradualism, a controversy that continues to this day (see later).

Finally, the non-European fossils from, e.g., Java, China, and South Africa could have been a strong stimulus for a wider view of human origins, but the prejudice of the leading British scientists crushed innovative hypotheses, and the critical skepticism on non-European roots of humankind due to eurocentristic perspectives and the fatal misinterpretation of the Piltdown man resulted in a stagnation of paleoanthropological theories (Dennell 2001). For this reason, the period after Darwin till 1930 could aptly be named—as Dennell (2001 p 52) comments—the “Age of Prejudices” as European (and North American) perceptions of nonwhites is concerned. The same author goes on to say: “... it is depressing to realize how much of what was written on human origins before WWII was little more than prejudice masquerading as science, it is impossible to understand the study of human origins without reference to these prejudices” (Dennell 2001 p 51). The reader is strongly recommended to read Dennell’s review (see also Hoßfeld 2005b).

1.4 Paleanthropology and politics: three symptomatic events

The misapplication of the Darwinian biological theory known as social Darwinism has been intensively discussed in another context (Mühlmann 1968; Zmarzlik 1969; Altnner 1981a, b; Müller-Hill 1988; Hofstadter 1995; Hawkins 1997; Dickens 2000; Junker 2004; Hoßfeld 2005b). It is well known that this biological ideology became highly influential after the World War I and especially during the period of National Socialism in Germany. That paleoanthropology was no free zone of research stems from different reasons: first, because paleoanthropology was a subject of the highly politically involved anthropology; second, because racial thinking and racist theory were interwoven with all aspects of daily life (Bowler 1976; Stepan 1982; Bowler 1986, 1988, 1996, 1997, 2001; Vogel 2001). The following three symptomatic cases are discussed to exemplify pars pro toto and exemplarily the impact of fascistic ideology on paleoanthropology during the “Third Reich” in Germany.

First, that archiological and paleoanthropological research was not spared by the ideology of the national socialists is documented, e.g., by the fact that the Neandertal-Museum, which had been inaugurated on May 1, 1937, was closed by order of the Reichsleiter für Vorgeschichte, Reinertz, on March 3, 1938. The arguments were that the Museum, which should demonstrate “... Deutsche Urgeschichte, soweit sie mit dem Neandertal in engster Beziehung steht,” did not fulfil the evaluation of an NS-commission (Beckmann 1987).
Second, even more inimical to the development of paleoanthropology during the Nazi era was the fact that the most outstanding German paleoanthropologist at that time, Franz Weidenreich (1873–1948), lost his *venia legendi* at the Johann Wolfgang Goethe University due to the persecution of the Jewish population. His emigration to the USA was an irreplaceable loss for anthropology in Germany. Since he was involved in research on the largest *H. erectus* sample from a single locality, Zhoukoudian, paleoanthropology in Germany lost a vital contact with the international scientific community. Thanks to Weidenreich (1943), there are brilliant documentations, casts, and descriptions of the “*Sinanthropus*”—skulls which were lost during an attempt to ship them to the USA (Shapiro 1974), and—equally important—he is viewed as the founder of the “multiregional theory of human evolution” (Wolpoff and Caspari 1997; Wolpoff 1999).

Third, it must be mentioned that during World War II, the first edition of Gerhard Heberer’s compendium “*Die Evolution der Organismen*” appeared (Heberer 1943). The anthropogenetical contributions deal with comparative anatomy (v. Krogh), paleontology (Gieseler), ethology, and archeology (Weinert). In the view of the scientific historians Hoßfeld and Junker (2003) and Hoßfeld (2005b), the chapter authored by Otto Reche is the only one that takes the model of synthetic Darwinism into account. There is full agreement with the following overall assessment: “...[es] handelt sich über weite Strecken um vergleichende Untersuchungen zur Stammesgeschichte der Menschheit, wie sie bereits im 19. Jahrhundert angestellt wurden, ergänzt durch neuere Daten aus Serologie und Paläontologie” (Hoßfeld and Junker 2003 p 107). The biological impact on paleoanthropology was minor, as long as in practice fossil discoveries came into the literature through collaboration between the archeologists who excavated them and the anatomists who described them. Foley (2001 p 6) states: “Archeologists and professors of anatomy seldom made a rich cocktail of Darwinian theory.” In general, the descriptive casuistic and more or less narrative approaches in paleoanthropology hold true for the period under discussion and for early postwar times, in Germany as well as in other countries (Tattersall 2000a; Foley 2001). The period leading up to World War II had seen the emergence of the evolutionary synthesis, but not until the 1950s did the innovative biological principles and methods begin to inform the paleoanthropological science too (Delisle 1995; Hoßfeld 1997, 2005b; Jahn 2000; Tattersall 2000a; Foley 2001; Corbey and Roebroeks 2001b; Hoßfeld and Junker 2003; Junker 2004) (Figure 1.6).

The selected examples given here demonstrate that social and political reasons far removed from paleoanthropology are important. Dennell (2001 p 45) is right when remembering “... that paradigmatic shifts in palaeoanthropology occur in response to a wider world, and are not wholly dependent on
internal evidence or individual personalities alone.” As Dennell demonstrated very convincingly, there are strong correlations of the shift of paleoanthropological interest from Asia to Africa, first with the transition from colonialism to independence in both Asia and Africa, second with the replacement of Europe by the USA as the dominant influence in paleoanthropology, and finally with the change from paleoanthropological narration to theory-guided analysis and explanation of behavioral changes and adaptations in the past.

1.5 Diversification of anthropology in the second half of the twentieth century

1.5.1 Early post-World War II period

The principal architects of the evolutionary synthesis mentioned previously gained more and more influence in World War II and postwar times, and their ideas shaped the field of biological anthropology including paleoanthropology (Tattersall 2000a; Foley 2001; Junker 2004; Hoßfeld 2005b). Although the fossil record had steadily increased, it was still a challenge to reconstruct the raw outlines of human evolution from the tiny catalog of human skeletal remains. There was tremendous progress as the hominin status (at that time taxonomically
tered hominid) of the small-brained, bipedal australopithecines became accepted due to the new results from fossils from Kromdraai Cave, excavated by Robert Broom in 1938, and further remains from Sterkfontein and Makapansgat, which were dug up in 1947. The diversity of the australopithecine taxa became further obvious from the megadontic *Paranthropus* from Swartkrans. The phylogenetic tree of our origins seemed to be quite simple. However, as the dating of known specimens was limited to the relative timescale “older than/younger than,” and as the techniques of absolute dating were unknown, there was a lot of uncertainty in the calibrations. Still the overall picture was a stepwise adaptive evolution from primitiveness to high-order hominids/hominins with modern shaped bodies and cultural skills. The evolutionary ladder reached from the gracile and robust australopithecines, i.e., very ape-like species, to the *Pithecanthropus* and *Sinanthropus* species. These had been lumped with the European *H. heidelbergensis* in a single species, *H. erectus*. This simple gradualistic model regarded the Middle Pleistocene “Java man,” “Peking man,” and the “Mauer man” as intermediate stages between the early taxa from Africa and the later Neanderthals, a well-documented fossil human group from late Ice Age sites in Europe and western Asia. Little wonder that in this confusion paleoanthropologists welcomed Mayr’s and Dobzhansky’s simplifying message that only one kind of hominid (new systematic: hominin) could have existed at any time and that virtually all developments in human evolution since Java man had taken place within the single, albeit variable, species *H. sapiens*. Tattersall (2000a p 3) suggests that Dobzhansky was influenced “by his newly arrived New York neighbor Franz Weidenreich, nowadays hailed as the father of ‘multiregional continuity’.”

Mayr took the position that at the most three successive species could be discerned within the genus *Homo*: *H. transvaalensis* (the australopithecines), *H. erectus*, and *H. sapiens* (including the Neanderthals). Few paleoanthropologists followed these pronouncements in all their details; however schemes of hominin evolution routinely came to incorporate the synthesis’ basic assumptions, whereby evolutionary change consisted simply of the gradual modification of lineages, usually no more than one, over long spans of time. Human evolution thus became the story of a long, single-minded struggle from primitiveness to perfection (Tattersall 2000a pp 1-16; contra Foley 2000). Successively there arose different problems from Mayr’s simplified model:

First, the evolutionary model was in contradiction to the mosaic pattern of the erroneously ancient, heavily encephalized Piltdown man. This problem disintegrated when in 1953 Piltdown was declared a hoax by authorities at the British Natural History Museum (Spencer 1990b). Second, the credibility of paleoanthropological results was severely flawed due to arbitrary taxonomic approaches. For this reason, there was a genuine need for the revision of the
overabundance of species and genus names, which had been applied liberally more or less without a taxonomical concept to hominid fossils. The taxonomic revision of the hominid/hominin sample asked for a strict taxonomic concept (Washburn 1953; Waegele 2000; Wiesemüller et al. 2003). Third, the uncertainty of the temporal allocation of fossils still reflected the absence of absolute dating techniques (Bishop and Miller 1972). In sum, it became more and more obvious that there was tremendous need for a new and precise strategy of physical anthropology within the frame of a revised synthetic theory of evolution (Washburn 1953; Cartmill 1990; Bowler 1997).

In the immediate postwar period, there arose a body of theory that swept away a host of conflicting notions about the nature of the evolutionary process. The elegantly simple concept stated that all evolutionary phenomena could be ascribed to a single mechanism: the gradual change of genes and gene frequencies within lineages of organisms under the guiding hand of natural selection (Cartmill 1990). The innovative population genetic approach was an important one for the field of paleoanthropology too as this was the time when the discipline absorbed the theoretical underpinnings that continue to influence it to the present day (Wuketits 1978; Vogel 1983; Foley 1987; Martin 1990; Jones et al. 1992; Delisle 1995; Tattersall 1995, 2000a, b; Foley 2001; Ruse 2005).

### 1.5.2 New strategies, concepts, and challenges in physical anthropology

In the early 1950s, there started an intensive discussion on the strategy of physical anthropology, i.e., the body of scientific theory and techniques with which it attacks its problems. Washburn (1953) was one of the protagonists of a new conception which was formulated in the new version demonstrated in Table 1.2. The main step was to reduce the speculative and narrative part of physical anthropology in favor of a thorough hypothesis testing. Cartmill (1990 p 189) puts it this way: “No doubt, people are different from the apes; but it is our job as scientists to explain those differences. Explanation, as opposed to mere storytelling, has to invoke law-like regularities connecting causes and effects.” This then was the challenge facing the evolutionary biological sciences (Popper 1968). Of greatest importance for the renewal of physical anthropology were the rejection of typological concepts and an increased interrelationship between the different parts of anthropology. The change from descriptive studies to the investigation of process and behavior brought about the integration of the problems of human evolution in the vast scientific field of mammalian evolutionary biology. Solutions to problems consequently followed, on the one hand
from paleontology, primatology, genetics, population genetics, and diverse medical sciences and on the other hand from the study of archeology and ethnology.

The great challenge of human evolutionary biology became the cultural factor as adaptations, human migrations, mating systems, population densities, diseases, and human ecology all became factors seen as essential to the explanation of our special human way of life (Washburn 1953; Vogel 1966; Osche 1983; Foley 1987). Washburn’s (1953 p 726) prognosis was: “If we would understand the process of human evolution, we need a modern dynamic biology and a deep appreciation of the history and functioning of culture. It is this necessity which gives all anthropology unity as a science.”

The recognizable post–World War II shape of paleoanthropology resulted from the belated acceptance of neo-Darwinian principles of evolutionary biology,
which were successively constituted in the 1930s and 1940s and that unified evolutionary biology under a single roof, sweeping away a huge package of mythological narrative thinking (Henke and Rothe 2006). Delisle (1995 p 217) suggested: “… that the evolutionary synthesis directly influenced on human paleontology [during the decade 1950–1960] every day practitioners in human paleontology almost solely through the general concepts and methods of the new systematics. Instead of being only a common core shared by a host of disciplines, the evolutionary synthesis should also be defined by the extent to which that core has been guiding current research in any one field.”

But, was the Modern Synthesis a real step forward in the right direction? Tattersall (2000a p 2) judges this step from an overcritical point of view and comments: “Sadly, however, the Synthesis was doomed to harden, much like a religion, into a dogma: a dogma whose heavy hand continues to oppress the science of human origins a half-century later.” He gives many arguments for this judgment, e.g., Dobzhansky’s lumping of the fossil hominids, with the conclusion “… that there existed no more than a single hominid species at any one time level” (Dobzhansky 1944 pp 261–262); further there was Mayr’s claim that humans did not speciate (Mayr 1950), a position which opened for later proponents the so-called “single-species-hypothesis.” Finally there came the controversy on “stasis” versus “punctuated equilibria” (Eldredge and Gould 1972). Tattersall (2000a p 5) complains that paleoanthropology was laggardly: “Slow to absorb the principles of the Synthesis, palaeoanthropology has been equally slow to augment these principles with recognition of the multifarious complexities of the evolutionary process.”

Is paleoanthropology really a discipline apart from the mainstream of biological thinking, and has the Modern Synthesis really shadowed the scientific work of paleoanthropologists? During the last 50 years, the last vestiges of the so-called “step ladder” model have been successfully refuted. While the unilinear and anagenetic models vanished from discussion, it became increasingly obvious that the process of human evolution is convincingly shown with multiple species, cladogenesis, and adaptive radiations, as well as the mechanisms and the process of the punctuated equilibrium model. Our current perspective is based on the broad acceptance of multiple species (slogan: “no tree but bushes”) and complexity of hominin diversity form (Cartmill 1990; Henke and Rothe 1994, 1999b, 2003).

There are reproaches against the Modern Synthesis as an obsolete theory which was uncritically used to give support for unilinealism, but not all recent protagonists of evolutionary thinking agree with the dichotomous equation “synthesis = unilinealism” and “multiple species = macroevolution.” Foley’s opposing arguments center on the point that paleoanthropology remained “… in fact […] blithely innocent of most theoretical issues” (Foley 2001 p 5). I incline
toward Foley’s (2005) opinion that the anagenetic models of the 1960s and 1970s owed, firstly, more to presynthesis notions of progress than to a model of adaptive fine-tuning and secondly, that Weidenreich’s position often reflected orthogenetic thinking rather than neo-Darwinian models of evolution (Foley 2001 p 6).

Concerning the first point, I would cite Vogel’s (1983 p 225) biological perspectives of anthropology and the so-called theory-deficit of the physical anthropology in Germany. He quotes Washburn (1953), who pleaded that the “... application of a constituent, experimentally verified, evolutionary theory is the first task of the physical anthropologist.” One should remember that Vogel’s paper was presented in 1981, which means that paleoanthropology and in a wider sense physical anthropology remained behind to other evolutionary sciences (Spiegel-Rösing and Schwidetzky 1982). Although paleoanthropology was not the foremost field of anthropological research in Germany, there appeared new editions of Die Evolution der Organismen (Heberer 1959, 1967–1974) and the textbook Menschliche Abstammungslehre (Heberer 1965a, b). In spite of these and other prominent publications in Germany in the 1960s and 1970s (Kurth 1962, 1968; Heberer 1968b; Hofer and Altner 1972; Kurth and Eibl-Eibesfeldt 1975; overviews in Hoßfeld 1997, 2005; Henke and Rothe 2006), there remained a theoretical vacuum in anthropology, and anthropology seemed here—as well as in other European countries—much more fossil-driven than hypothesis-guided. Cartmill (1990, p173) claimed in the early 1990s: “Paleoanthropology should aim at increasing its theoretical content by reducing the list of qualitative human uniquenesses—and eliminating it altogether if possible.” The evaluation of the scientific reasons that perpetuated inadequate evolutionary approaches in paleoanthropology and caused the complacency with insufficient models is still incompleted. Whether “... the Synthesis was doomed to harden much like a religion, into dogma as Tattersall claims (2000a p 2), contra Foley (2001), is a continuing issue; but most can agree with the statement that paleoanthropology was not explicitly theoretical but was descriptive and in part excessively narrative (Bowler 1996, 2001; Foley 2001; Ickerodt 2005).

In contrast to the European situation, paleoanthropologists in the USA have traditionally been trained as physical and cultural anthropologists. This caused a different approach to the research in human evolution; in the traditional European system as a rule archeologists dug up the fossils and they themselves or anatomists—not physical anthropologists—described them and perpetuated the problem. The simplicity of the narrative approaches also hampered progress. The idea of human uniqueness was emphasized (Cartmill 1990) and hominin fossils were treated casuistically. Foley (2001 p 7) confesses that it was “... frustration with the combination of an absence of evolutionary theory in human evolution and assumptions about human uniqueness that led me to

Historical overview of paleoanthropological research
His sophisticated approach tries to explain the process of human evolution and the human adaptive strategy as intersection of the biological categories to which hominins (in Foley’s original hominids) belong (Figure 1.7). Although most hominin diversity can be explained by evolutionary changes caused by geographical–climatological factors,

there is a vast explanatory field within the life sciences, particularly comparative primatology, sociobiology, paleoecology, and paleogenetics. The multidisciplinarity of paleoanthropology is demonstrated in Figure 1.8, however, without a detailed special ranking concerning the importance of the cooperating disciplines. The preference for cooperation naturally depends on the paleoanthropological
problems that have to be solved, but it should be remembered that inter-, multi-, and transdisciplinarity is a learning process and an everlasting challenge for science (Mittelstraß 1989; Drilling 1992; Eggert 1995; Porr 1998; Henke 2006).

The last half-century has witnessed a dramatic improvement in our understanding of the process of human evolution, due to new approaches and techniques as well as a tremendous increase in the fossil record (Andrews and Franzen 1984; Franzen 1994; Henke and Rothe 1994; Ullrich 1995, 1999; Johanson and Edgar 1996; Hartwig 2002; Schwartz and Tattersall 2002, 2003). Profound knowledge of the relationship between form and function has come from innovations in mechanical engineering, light microscopy and REM, 2D- and 3D-tomography as well as from multivariate statistics (Grupe and Peters 2003; Zollikofer and Ponce de Leon 2005). Furthermore, evolutionary and developmental morphology (Mingh-Purvis and McNamara 2002) and physiology (Martin 1990) has contributed to a better understanding of form–function complexes (Ciochon and Corruccini 1984; Oxnard 1984; Aiello and Dean 1990; Anapol et al. 2004; Ross and Kay 2004). Hennig’s Phylogenetic Systematics (Hennig 1966, 1982, 1984), which was first published in German in 1950 without gaining much attention (Hennig 1950), revolutionized phylogenetic discussion in concert with tremendously increased skills in taxonomy and computer techniques (Rieppel 1999; Waegele 2000; Wiesmüller et al. 2003).
With the improvement of absolute dating techniques (e.g., radiocarbon and other isotopic calibrations) and in relative dating by faunal complexes, the chronological pattern of human evolution was more precisely observed, and it became seen through the application of “molecular clocks” that the branching of the hominin line coincided with the aridification of the East African Rift Valley (Bishop and Miller 1972; Howell 1978; Magori et al. 1996; Bromage and Schrenk 1999).

The more that research on human evolution concentrated on the African continent, the more successful those paleoanthropologists with a licence to dig became, especially as they extended their campaigns to Miocene as well as to Pli- and Pleistocene strata. Leakey’s luck (Cole 1975; Isaac and McCown 1976) at Olduvai Gorge—which had been discovered by the German neurologist Wilhelm Kattwinkel in 1911 (Glowatzki 1979) and was successfully explored for the first time in 1913 by Reck (1925), as well as that of his family members in Koobi Fora and diverse other East African sites—has resulted from tremendous efforts (Leakey and Leakey 1978; Grine 1988; Wood 1991; Tobias 1991; Walker and Leakey 1993). Beside the activities of the Leakey family, there should be mentioned the successful expeditions of Francis Clark Howell in Omo, Glynn Isaac in Olorgesailie, and of course the famous Afar Research Expedition (Johanson and Edey 1980; Johanson and Edgar 1996). Finally, the Hominid Corridor Research Project of Timothy Bromage and Friedemann Schrenk in Malawi (Bromage and Schrenk 1999) must be alluded to as well as the activities of Brigitte Senut and Michael Pickford in Tansania (Pickford and Senut 2001) and Brunet et al. (2002) in Chad. Africa has become the “Mecca” of paleoanthropologists. There is little doubt now that Africa was the “cradle of mankind,” just as Darwin proposed.

In spite of the fact that the “fossil hunting” has mostly been done in Africa, there are many activities in other parts of the Old World allowing us to learn more about the pattern of hominid migration and development. Exciting new fossils and findings from, e.g., Atapuerca (Spain) (Arsuaga et al. 1999), Apidima (Greece) (Pitsios 1999), Ceprano (Italy) (Ascenzi et al. 2000), Schöningen (Thieme 1996), Dmanisi (Georgia) (Bräuer et al. 1995; Henke et al. 1995; Gabunia et al. 1999, 2002; Vekua et al. 2002) and many non-European sites (Delson 1985; Rightmire 1990; Franzen 1994; Johanson and Edgar 1996; Delson et al. 2000; Brunet et al. 2002; Schwartz and Tattersall 2003), and the astonishing fossils from Flores (Indonesia) (Brown et al. 2004) demonstrate that paleoanthropology is a field of research with never-ending surprises and new perspectives. No doubt, the half-life of our theoretical models is very short, but there has undoubtedly been enormous progress. Beside the paleontological fieldwork, which has also produced valuable data for the reconstruction of the ecological
niches of our ancestors (Bromage and Schrenk 1999), the research field of primatology has become increasingly important for anthropological modeling (Martin 1990; MacPhee 1993; Fleagle 1999; Groves 2001). The paradigm of behavioral ecology and sociobiology has shaped our hypotheses on food choice, foraging patterns and food detection, as well as on food sharing and intra- and intergroup relations. Studies on the evolution of social behavioral systems, kin selection, intersexual and intrasexual selection, cognitive abilities, tool using and tool making, Machiavellian strategies, competition, coalitions and alliances, that is on the total complexity of social systems in primates, especially in apes, have become essential for paleoanthropological modeling (Henke and Rothe 2003; Henke 2003a; Rothe and Henke 2005). Beside the primatological field studies, which have given us a totally new view on the cultural capabilities of nonhuman primates, there is much to learn about our brains and the development of language from all kinds of laboratory research [e.g., molecular biology: O’Rourke et al. (2000), Enard (2005), psychobiology: Tomasello (1999)].

Still, the “old questions” in paleoanthropology (v. Koenigswald 1958) remain unsolved and valid. One hundred and fifty years after the excavation of the name-giving fossils from the Neander Valley, near Düsseldorf in Germany, the “role of the Neanderthals” is the subject of intense discussion (Spencer 1984; Stringer and Gamble 1993; Henke and Rothe 1994, 1999b; Tattersall 1995; Krings et al. 1997; Wolpoff 1999; Relethford 2001; Finlayson 2004). Also debated are the Out of Africa models (Bräuer 1984; Bräuer and Smith 1992; Wolpoff 1999; Relethford 2001; Finlayson 2004). Maybe the humorous definition “humans are animals who wonder intensively and endlessly about their origin” is the most appropriate one for our species. We are in an age of tremendous progress in paleoanthropological research and knowledge, which hopefully is evident in the contributions of this handbook.

1.6 Final remarks

The preceding survey demonstrates that, compared with other biological sciences, progress in paleoanthropology was slow for most of the last century. The reason for this is multifaceted and still under discussion (Tattersall 2000a; Dennell 2001; Foley 2001; Corbey and Roebroeks 2001b). One important problem leading to slow progress was a descriptive and nonanalytical approach due to theoretical deficits. Innovative approaches and the positioning of paleoanthropology within a complex research strategy of evolutionary biology integrating all facets of comparative primatology, behavior ecology, sociobiology, molecular and
population genetics (Martin 1990; Jones et al. 1992; Jobling et al. 2004), as well as archeology (Klein 1989; Gamble 1999; Klein and Edgar 2002), archeometry (Herrmann 1986, 1994), and earth sciences (Isaac and McKown 1976; Foley 1987; Bromage and Schrenk 1999) have led to a change. Within a broadly based scientific field and a consequently theory-guided empirical approach, paleoanthropology has become during the last decades a “mature” science. The integration of paleoanthropology into the wider set of biological sciences and the system theory of evolution is the road map paleoanthropologists should follow to achieve reasonable and credible scenarios of the past. Paleoanthropologists should be aware of the principal aims of scientific work (Popper 1983; Mahner and Bunge 2000) to avoid the pitfalls of a narrative, storytelling, as well as fossil- and journalism-driven science (Cartmill 1990; Tattersall 1995; Franck 1997, 1998a; White 2000; Schrenk and Bromage 2002).

With the principal and methodological suppositions of the biological and neighboring sciences in mind, paleoanthropologists should be aware of science history too, or, in the words of Theunissen (2001 p 147): “That ‘history matters’ is perfectly obvious if it is taken to mean that the present cannot be understood without reference to the past.” Although this statement seems to be trivial, there is a dilemma which Foley (1987) mentioned: there is no doubt that the past was a “foreign country” and that our forerunners did things differently there. The challenge of paleoanthropology is to improve our scientific methodology, to learn more about us and our origin, to reach a biological and sociobiological self-conception (Vogel 2000). The historicist approach may give rise to reflexive doubt. But Dennell (2001 p 64) is right when he says: “Whilst we might like to think that palaeoanthropology is a discipline that unifies humanity and helps combat racial, sexist, and other types of prejudice simply because of its focus on the origins of humankind, we should not forget it did the opposite for at least the first half of the 20th century.”

The confidence of being on the right road raises the risk of complacency. And herein lies the essential argument that history really matters, that one should never be too comfortable for too long a time with an idea. What is really needed is a real and intensive dialogue between the many disciplines, which cooperate to unravel the process of human origin. Returning to the introductory question of whether science history matters within this canon of disciplines, the conclusion must be that history forms a constitutive part of present day research. The history of paleoanthropology teaches us, in all facets and details, that the reconstruction of our origin is a real challenge; McHenry (1996 p 86) puts it: “One needs to make the best of our tiny sample of life in the past, to be open to new discoveries and ideas, and to enjoy the pleasure of learning and changing.”
References

Abel O (1931) Die Stellung des Menschen im Rahmen der Wirbeltiere. Gustav Fischer Verlag, Jena


Bishop WW, Miller JA (eds) (1972) Calibration of hominoid evolution. Recent advances in isotopic and other dating methods applicable to the origin of man. Scottish Academic Press, Edinburgh

Bolk L (1926) Das Problem der Menschwerdung. G. Fischer-Verlag, Jena


Corbey R, Roebroeks W (2001b) Studying human origins. Amsterdam University Press, Amsterdam


origins. Amsterdam University Press, Amsterdam, pp 107–122
Desmond A (1997) Huxley: From Devil’s disciple to evolution’s high priest. Addison Wesley, Reading, Mass
Dilthey W (1883) Einleitung in die Geisteswissenschaften. o.A
Franck G (1998a) Ökonomie der Aufmerksamkeit. Hanser, München
geological setting, and age. Science 288: 1019–1025
Gorjanović‐Kramberger D (1906) Der diluviale Mensch von Krapina in Kroatia. Ein Beitrag zur Paläoanthropologie. C.W. Kreidels Verlag, Wiesbaden
Grimm H (1961) Einführung in die Anthropologie. VEB Gustav Fischer Verlag, Jena
Haeckel E (1866) Generelle Morphologie der Organismen. 2 Bände. G. Reimer Verlag, Berlin
Haeckel E (1899) Die neuen Fossilien des Mesozoikums. 4 Bände. G. Reimer Verlag, Berlin
Heberer G (Hrsg.) (1968b) Der gerechtfertigte Haeckel. Einblicke in seine Schriften aus Anlaß des Erscheinens des Hauptwerkes “Generelle Morphologie der Organismen” vor 100 Jahren. G. Fischer Verlag, Stuttgart
Herrmann B (Hrsg.) (1994) Archäometrie. Naturwissenschaftliche Analyse von...
Sachüberresten. Springer-Verlag, Berlin Heidelberg New York
Hooton EA (1931) Up from the ape. The Macmillan Company, New York
Huxley TH (1863) Evidences as to man’s place in nature. Williams and Norgate, London
Historical overview of paleoanthropological research

Kollmann J (1885) Das Ueberwintern von europäischen Frosch- und Tritonlarven und die Umwandlung des mexikanischen Axolotl. Verhandlungen der Naturforschenden Gesellschaft in Basel 7: 387–398
Kuhn TS (1962) Die Struktur wissenschaftlicher Revolutionen. Suhrkamp, Frankfurt am Main
Le Gros Clark WE (1934) Early forerunners of man. Baillière, Tindall and Cox, London

Menke P (this volume) The ontogeny–phylogeny nexus in the nutshell: Implications for primatology and paleoanthropology


Mühlmann WE (1968) Geschichte der Anthropologie. 2. Auflage, Athenaum Verlag, Frankfurt a.M., Bonn


Popper K (1959b) The logic of scientific discovery. University of Toronto Press, Toronto


Reche O (1937) Begrüßung, Verhandlungen der Deutschen Gesellschaft für Physische Anthropologie 8: 1–3


Rieppel O (1999) Einführung in die computergestützte Kladistik. Verlag Dr. Friedrich Pfeil, München


Schmerling C-P (1833) Recherches sur les ossements fossiles découverts dans les caverns de la Province de Liège. P-J Collardin, Liège
Schwalbe G (1906) Studien zur Vorgeschichte des Menschen. E. Schweizerbart’sche Verlagsbuchhandlung (E. Nägele), Stuttgart
Starck D (1962) Der heutige Stand des Fetalisationsproblems. Verlag Paul Parey, Hamburg Berlin
Tattersall I (2000b) Once we were not alone. Sci Am 2: 56–63
Historical overview of paleoanthropological research

Werth E (1921) Der fossile Mensch. Grundzüge einer Paläoanthropologie. Verlag Gebrüder Borntraeger, Berlin
2 Evolutionary Theory in Philosophical Focus

Philippe Huneman

Abstract

This chapter surveys the philosophical problems raised by the two Darwinian claims of the existence of Tree of Life and the explanatory power of natural selection. It explores the specificity of explanations by natural selection, emphasizing the high context dependency of any process of selection. Some consequences are drawn about the difficulty of those explanations to fit a nomological model of explanation, and the irreducibility of their historic-narrative dimension. The paper introduces debates about units of selection, stating the compelling force of genic selectionism but highlighting some critiques. It then addresses the limitations of selectionist explanations: the compared status of selection, drift, and phylogenetic inertia are investigated, and the debates over adaptationism are presented, with the aim of defining the varieties of adaptationisms as research programs. In order to assess the scope of natural selection, the chapter addresses weak and strong challenges to the Synthetic theory of evolution both from paleontology (punctuated equilibria, Gould’s contingency thesis) and the evolutionary theory of development. We finally sketch some consequences of evolutionary theory concerning philosophical questions about human nature, on the basis of the hypothesis of the universality of selectionist explanations: this part deals mostly with epistemology and psychology.

2.1 Introduction

The theory of evolution, from Darwin to the Modern Synthesis formulation, provided a framework of explanatory strategies to explain diversity and adaptation in the living realm. Considered on a large scale, Darwinian science advanced and justified two main claims: the Tree of Life, meaning that all the extant living species are always historical results of common descent, and the Selection hypothesis, meaning that one of the most important mechanisms to account for those transformations is “natural selection.” Hence, it added to the ancient life sciences
a new *explanandum*, e.g., phylogenesis, and a new *explanans*\(^1\)—natural selection—which is also an explanatory resource for more traditional kinds of problems.

Of course, the consequences of the two main Darwinian claims were not recognized immediately; people were too much concerned by the two metaphysical issues of evolution versus creationism, and of the animal origins of man. It took a little less than a century to acquire the historical distance that enables one to rightly appreciate the novelty of Darwinism and this happened with the Modern Synthesis. For the Synthesis, population genetics has a central status within evolutionary thinking: historically, two of the founders of the synthesis, Fisher and Wright, were population geneticists, and some fundamental statements of evolutionary biology are enunciated in population genetics (Fisher’s theorem, Hardy–Weinberg equilibrium, etc.); conceptually, the definition of evolution, as a change of the gene frequencies in the gene pool, lies in the field of population genetics. This essential feature of the theory was not conceivable in the time of the first Darwinians, since the gradualist view of transformism seemed to contradict the discontinuous vision of organisms as mosaics of traits that Mendelian genetics had presupposed. It is often and truly said that neo-Darwinism unified Darwin and Mendel, thereby superseding such an apparent conflict.\(^2\) Weissmann, by separating soma and germen and advocating that there was no transmission of the acquired characters, gave a clear meaning to the difference between Darwinism and Lamarckism and allowed his followers to regard only what is in the germen as the substrate of evolution, enabling the future integration of genetics within evolutionary biology. Moreover, Weissmann proved impossible the theories of heredity and variation maintained by many biologists and Darwin himself, according to which hereditary traits could be produced within the individual organism’s cells and flow continuously from them. It then became possible for geneticists to propose mechanisms of heredity and variation in which the Darwinian theory of natural selection had only to assess the facts of heredity and variation, without being in principle committed to any theory of heredity (even if that was what Darwin actually did).

In order to grasp the new kind of epistemological problems brought by the two Darwinian contentions, it is quite useful to recall the features of the earlier biology that they replaced. The main explanans of diversity and adaptation before

---

1. In the usual vocabulary of the philosophy of science, *explanandum* means what is to be explained, whereas *explanans* means what explains the *explanandum*.
2. For an account of the conceptual transformations which led from Darwin to neo-Darwinism through the successful synthesis of Mendelian genetics and Darwinian hypothesis, see Gayon (1998).
Darwin was, as we know, the divine design, although other hypotheses were being proposed more and more often, especially the evolutionary theory of Lamarck, which was adopted by Geoffroy Saint Hilaire and many morphologists at the beginning of the nineteenth century. This design was invoked to account for some *prima facie* teleological features of the living world, such as the fine adaptation of organisms to their environment, or the fine tuning of the mechanisms of biological function, or, in the end, the proportions of individuals in various species and the geographical relationships between species. The divine design yielded simultaneously the *individual designs* of organisms, unlikely to be produced by the mere laws of physics, and the *design of the whole nature* that Linnaeus called the “economy of nature.” The Selection hypothesis gave a powerful explanation of those two designs, since adaptations of organisms as well as distributions of species in a population were likely to be understood by appealing to the process of natural selection (even if other mechanisms like Lamarckian ones were also used by the first Darwinians\(^3\)). Since the result of such a process is a Tree of Life, biologists justify the striking similarity of forms between different species of the same genus, or even different genera of the same family—this fact being an immediate result of the common descent of different members of a same taxon.

However, the rise of Darwinism did not mean a total shift of the relevant questions and tools in biology. Rather than deleting centuries of research in the science of life, Darwinism gave a new and coherent meaning to some admitted facts and descriptions. Instead of rejecting teleology outside science, it provided a way of interpreting teleological phenomena so that they did not depend on nonnaturalistic presuppositions, such as hidden intentions of the organisms or their creator; it kept the result of the traditional taxonomist’s effort and conceived the systematic proximities in the classification of species as historical affiliations, as Darwin himself noticed at the end of the *Origin of Species* (even if, of course, the Darwinian views raised new questions and permitted new criteria and methods for systematists (Ghiselin 1980)).

So, evolutionary theory appears to us as the most successful and integrative framework for research strategies in biology. Before investigating the details of the epistemological challenges raised by the two Darwinian claims, it is therefore useful to situate evolutionary theory within the whole of biology. Here, Ernst Mayr’s conception of explanation in life sciences will be of some help. In effect,

---

\(^3\) On the progressive extension of Darwin’s theory, and all the slight nuances that made it very different from what we use now and the sharp picture here drawn, see Ghiselin (1969), Ospovat (1981), Ghiselin (1969) and Bowler (1989). Ospovat emphasizes the conditions for Darwinism in the work of morphologists like Geoffroy and Owen.
Mayr distinguished two kinds of causes as different answers to the question “why” (Mayr 1961). When asked: “why does this bird fly along the seashore to the south?,” you can answer by pointing out its physiology, respiratory system, the diverse pressures on its wings, and the streams of air around it: this indicates the proximate causes of the bird’s flight. But you can also answer by emphasizing that the way it takes to go to the south curiously corresponds to the old demarcation of the continents, and you will understand that this is a result of natural selection acting on this species of bird to improve its time of migration. This is the ultimate cause of the bird’s trajectory. Notice that the first causes concern exclusively one bird, and each bird is concerned by them in the same way, meaning that they are generic causes. On the contrary, the ultimate causes collectively concern the ancestors of this bird and not the bird itself.

Notice also, and this will be of importance for all epistemological considerations, that the two kinds of causes do not answer exactly the same question: the former answers the question “why does the bird fly along the seashore (rather than being unable to fly)?”, whereas the latter answers “why does the bird fly along the seashore (rather than somewhere else)?” This distinction, highlighted in another context by Sober (1986), means that the two kinds of causes are embedded in different explanatory strategies. As Mayr would remark, a complete biological explanation of a phenomenon makes use of all those strategies. And the two kinds of causes correspond to two kinds of biological discipline: on the one hand, as sciences of the proximate causes we have molecular biology, physiology, endocrinology, etc. while on the other hand, as sciences of the ultimate causes we have all those disciplines belonging to evolutionary biology: population genetics, ecology, paleontology, etc.

Having characterized evolutionary theory as a specific set of research programs within biology, and those programs being defined by the use of the hypothesis of natural selection, we can present some evolutionary problems raised by evolutionary theory. These will concern essentially the nature and the limits of the explanation by natural selection. Thus, I will reveal these two kinds of problems by picking out, in each category, one or two fundamental and currently debated issues. Then I will stress some important consequences of evolutionary biology upon philosophical theorizing about human nature.

We must nevertheless notice that all those issues involve both biological and philosophical considerations. They will sometimes be closer to theoretical biology and the methodology of biology than to philosophy but will sometimes include apparently pure matters of metaphysics that make no difference in empirical science. However, I claim that there is a set of problems raised by evolutionary theory that is of essential interest for philosophy, but that cannot be handled by the traditional means of a general philosophy of science—and that
therefore must constantly appeal to considerations of theoretical biology. The fact that the Modern Synthesis has a unique character compared to other improvements in science (Shapere 1980), is surely one of the reasons for this peculiar status of the philosophical problems raised by evolutionary theory. However, given this special status, authors contributing to the debates are either philosophers of science, like Hull, Sober, Rosenberg, or Kitcher, or sometimes biologists who may have made major contributions to evolutionary biology, like Mayr, Gould, Maynard-Smith, Williams, or Lewontin. Philosophy of biology partly emerged from the dissatisfaction of philosophers of science with the logical positivistic program and their will to find new paths toward unsolved questions, and partly from the need, felt by biologists, of conceptual elucidations of the bases of their practice and of the consequences of their theories.

2.2 Evolutionary biology: its challenges for philosophy of science

2.2.1 What is selectionist explanation?

2.2.1.1 The process of selection and the property of fitness

Natural selection is a process that is expected to take place whenever the following requirements are fulfilled: there is a set of individuals; those individuals reproduce; there is variation among them and those variations are likely to be hereditarily transmitted; due to interaction with the environment, some varying properties provide their bearer with a chance to leave more offspring than individuals who lack such a property.

No matter what the entities are that fulfill those requirements, their set is, from now on, susceptible to be affected by natural selection. For this reason, people have proposed a theory of natural selection of macromolecules to account for the origins of life (Eigen 1983; Maynard-Smith and Szathmary 1995), or theories of natural selection of ideal elements in order to explain cultural evolution (Cavalli-Sforza and Feldman 1981; Boyd and Richerson 1985; Campbell 1990). The methodological problem here is to invoke a process of heredity, which is not as obvious as in the case of genes.

And, reciprocally, when one meets a set of individuals fulfilling these requirements, one can assume that those individuals have undergone natural selection so that their properties are the effects of natural selection (or, more precisely, the fact that they have the properties that they have, and not some other properties, is the effect of selection).
Let us state some characteristics of explanation by natural selection. First, this belongs to what Mayr (1959b) called “population thinking,” e.g., the explanandum has to be or to belong to a class of entities—what we call a population—for otherwise the differential reproductive success that is the result of having or not having a property, which in turn is what is meant by the word “selected,” would not be determinable.

This explanation by selection might be contrasted with what Sober (1984) called “developmental explanation,” namely, an explanation of the property of an individual appealing to the process through which it was acquired. The developmental explanation of the composition of a football team is the sum of the experiences of each of its players; the selective explanation is the choice of the team by the manager, who set a criterion of competence and then evaluated all available football players by this criterion. What is peculiar to selectionist explanation is the fact that there are no determined criteria of admission other than reproductive success.

This brings us to an essential characteristic of selectionist explanation, namely the fact that it is always selection for effects; hence it is blind to causes. No matter whether a red deer reaches reproductive success over his congeners as a result of his higher race speed, or through his visual abilities to detect predators: in both cases, the fact that he leaves more offspring will mean that his genes (among them, the ones bound to the decisive property) will be more greatly represented in the following generation. This sole fact is basic to natural selection and to selectionist explanations. According to a distinction made by Sober (1984), selection of something X (for example the red deer) is always selection for some property A enhancing survival and reproductive success (but of course, other properties, linked with A in X, are also selected of).

If we measure the selective advantage conferred on an individual by its properties, and use the term “fitness” for such a measure, then different properties (different in physical and chemical terms) will be likely to have the same fitness (always measured in a given environment). Thus, two consequences follow concerning fitness. First, fitness is what philosophers of science call “supervenient” on the physical and chemical properties of traits. This means that, if two traits are different, they may have the same fitness or not, but if two traits have different fitnesses, they must be physically or chemically different. Supervenience, so defined, implies “multirealisability,” meaning that a same fitness can be realized by various ontologically different properties and devices (Rosenberg 1978; Brandon 1990; Sober 1993).

---

4 Here, there is no direct reference to the embryological development.
The second consequence is that this property of fitness, since it depends on population-thinking explanatory strategies, has to be thought of as a probabilistic one, hence as a “propension” (Mills and Beatty 1979; Brandon 1990\(^5\)). This has an obvious reason: fitness is indicated by differential reproductive success, hence by the number of offspring. But two different individuals can have the same fitness and nonetheless leave different numbers of offspring. Mills and Beatty used the example of the twins, sharing a same genotype in a same environment, hence having the same fitness; nevertheless, one of them is struck by the lightning while still a young man, whereas the other mates and has six children. Thus, the actual number of offspring cannot represent the fitness; but fitness has to be measured by the expected number of offspring, which is a probabilistic parameter. A given individual may not leave the number of offspring stated in its “fitness.” One can easily see here that, if this were the case, then the fittest individual would be the one who leaves the most offspring, and evolution as the “survival of the fittest” or the reproductive success of the fittest would be a tautology. So, the idea of fitness defined according to a propensionist theory allows us to avoid the charge of tautology recurrently raised against Darwinism.

If fitness is a supervenient property, this entails important consequences for the relationship between biology and the physical sciences. In short, no necessary physical statement can account for biological phenomena involving selection—therefore, fitness—since the same fitness could be realized by other physical matters of facts, laws, and properties. Hence, evolutionary theory supervenes on the physical propositions and theories. But what about the status of selectionist explanations in biology when compared to the explanations in physical sciences?

\subsection*{2.2.1.2 Laws and selectionist explanation}

Here we come across the rather entangled philosophical topic of scientific laws. Physicists state laws of nature. However interpreted, those laws of nature are general statements formulated in the modality of necessity.\(^6\) The usual puzzle in philosophy of science is to find a criterion distinguishing accidental generalities and laws (Ayer 1956). As an answer, Dretske (1977) claimed that laws have to be conceived as relationships between universals. But in any case, laws should support counterfactuals: this means that if some variables are changed within

---

\(^5\) Critique is made in Rosenberg (1982); other recent critiques led to the precision of the propensity definitions (Sober 2001; Ariew 2004).

\(^6\) I leave apart here the difference between empirical and a priori laws.
them, the results should be affected in a regular way. This implies that lawlike generalizations can be used in explanations, whereas accidental generalizations seem not to allow such a use and even less a predictive use. The positivistic account of science viewed explanation as a deductive argument whose conclusion is the explanans, and whose premise sets some laws of nature with some particular statements of facts [the so-called DN account of science: Hempel (1965)].

In a provocative chapter of his *Philosophy of Scientific Realism*, Smart claimed that there are no biological laws, since any law has to be reliable for any individual, e.g., has to be stated in the form “for any X, P(X).” But in evolutionary theory, we only have statements concerning limited sets of entities, like teleost fishes, or more generally, birds in America, or *Equus*. People can forge seemingly law-like general assertions on the basis of such generalizations, like Dollo’s law concerning irreversibility in evolution, or Cope’s law concerning the increasing of size in populations; but they are nevertheless still spatio-temporally situated general statements lacking any nomothetic necessity. Necessity is supposed to hold for any individual of a given kind with no specification of space and time. These regularities fail to explain but merely describe; they are not predictive since they always find exceptions. For Smart, such biological regularities are like the schemata of engineers and are in the same way embedded in laws of physics. Even the universality of genetic code is a generalization on our planet, due to the contingent reason of common ascendance (contingent regarding the code itself), since the same correspondence laws between nucleotides and amino acids are not to be expected on any planet. For Beatty (1997), this contingency of generalized propositions affects the whole of the supposed law-like statements in biology. So the DN account fails to represent evolutionary theory.

The only evolutionary statement that could be a law is, thus, the one enunciating the process of natural selection since it specifies no particular entity. Philosophers debate about the nomothetic status of this principle of natural selection (PNS) (Bock and Von Wahlert 1963; Sober 1984, 1997; Brandon 1996, 1997; Rosenberg 1985, 1994, 2001). Rosenberg argues that the PNS is the only law of biology and relies on Williams’ (1970) axiomatization of the theory, which conceives fitness as an undefined primitive term, e.g., a term which in some definitions, in some contexts, can be given only outside evolutionary theory, in

---

7 Taking a famous example from Goodman, what predictions could I infer from “all the men in this room are third sons” (unless I have some additive information on those men, like: “they are all attending a third sons’ meeting”, etc.)?

8 Rosenberg (1994) also denies that biology has laws in the sense of physical laws since it supervenes on all the physical laws and hence can only pick out disjunctions of laws applied in limited contexts.
another theory. But, even if by convention we say that it is a law, we still face the question of its differences from the other kinds of law.\textsuperscript{9} In effect, unlike physical laws, the PNS does not state \textit{any natural kind} of property such as mass, electric charge, etc. The only property involved in its formulation is fitness, which is a mere supervenient property.

So the PNS becomes the equivalent of a physical law—stated in probabilistic language, of course—only as soon as some physical characters of the properties contributing to fitness are \textit{specified}, a specification which is always context dependent.\textsuperscript{10} For instance, the “optimal shift towards viviparity” described by Williams (1966) in some marine fishes results from a kind of law, since he stated the parameters ruling the selection pressures (density of predators, physiological cost of reproduction); parameters which in turn determine a range of relevant physical properties for selection. Hence, in this case the schema becomes predictive, and we can test it by building experiments in which the values of the variables concerned vary. This idea, however, does not exhaust all biological regularities, principally the aforementioned ones found in paleontology. Thus, recalling the two claims of evolutionary theory concerning both the Pattern and the Process of evolution, this way of constructing law-like sentences through the PNS is mainly relevant to the Process of evolution, whereas the Pattern is most likely to show nonexplanatory regularities.

So, if evolutionary theory is not, as Smart contended, a nomothetic science, neither is it a class of empirical generalizations combined with some mathematical tools. Moreover, in addition to the PNS, there is surely a set of genuine laws in evolutionary theory, since its core, population genetics, provides some models, such as the Hardy–Weinberg equilibrium, which prescribes a nomological necessity to any pool of genes in an infinite population. However, those kinds of propositions are not so much empirical laws as \textit{mathematical laws}. They define a sort of mathematics of genes. Such models are in no case a description of any actual population, for in order to be applied to populations they have to integrate empirical content, i.e., by fixing the fitness coefficients of alleles. But this is not the same thing as fixing the parameters (mass, charge, etc.) in any standard

\textsuperscript{9} By changing the definition of what counts as law, and, more precisely, weakening the DN requisites on laws of nature, one can imagine that there is a continuum of kinds of laws instead of a sharp boundary between accidental and nomological generalization. For example, relying on the supporting counterfactual requisite, Woodward (2001) defines laws as statements invariant through a sort of change in the explanandum. This enables him to count several laws, like Mendel’s laws, in biology, and then account for the predictive and explanatory role of an accidentally general statement such as the universality of the genetic code. My point is that characterizing the status of the principle of natural selection within such a continuum is still at stake.

\textsuperscript{10} A similar position is upheld in Brandon (1996) and Michod (1981).
physical case, because fitness can only be \textit{locally} defined, its relevant parameters being determined by the environment considered.\footnote{For example, color of moths is a fitness parameter in industrial melanism only because there are predators capable of vision.} And, even worse, those parameters are likely to change without change of environment since many cases of selection are frequency dependent. Admittedly, over three decades, after Maynard-Smith (1982), we have developed a powerful mathematical tool to build models in cases in which selection is frequency dependent, e.g., the value of a trait in an individual depends on what other individuals are and do: this is \textit{evolutionary game theory}, which can provide models in which ordinary population genetics fails because it treats fitness as a property of individuals and hence cannot forge models when fitness depends on frequency. The status of those models, however, is the same as that of the classical models of population genetics. Maynard-Smith (1982) insisted on the fact that one has to investigate the \textit{strategy set} before applying any game theoretical model to empirical cases, which means that by itself, game theoretical theorems and proofs, no matter how illuminating, do not have empirical content. So, in a way, evolution contains both statements stronger than physical laws (since they are purely mathematical models) and statements nomothetically weaker such as those derived from the PNS by its empirical instantiation. Rather than a law, the PNS in the end proves to be an \textit{explanatory schema}, providing ways of explaining and building models through its more or less empirical instantiations. The least empirically instantiated are models of population genetics; at the most empirically instantiated level, we have law-like generalizations, such as paleontological ones. In a way, it is a matter of convention whether or not to call them “laws”: the point is just to determine their epistemological nature.

\subsection*{2.2.1.3 Historical narratives and selective mechanisms}

As Bock and Von Wahlert (1963) wrote, one must distinguish the processes of evolution, which involve—but are not to be equated with—natural selection, and the outcome of evolution, namely phylogenies and the Tree of Life. However, no actual process of evolution could be understood solely through knowledge of mechanisms and without historical data. Let us give an example. Many terrestrial vertebrates are tetrapods. One could imagine a selective hypothesis concerning the adaptive origins of their four limbs since they are obviously adaptive for locomotion. However, there is another reason for those four limbs: marine ancestors of those vertebrates had four fins, so the four limbs are a legacy,
resulting from what we can call “phylogenetic inertia.” The point is that natural selection often explains the appearance of such traits but not in this precise clade, so the selectionist explanation has to be historically situated in order to determine what the correct explanandum is, for which natural selection would be the right explanans. Thus, no understanding of the presence of characters in the organisms of a given population or species is available through the sole application of models of natural selection. The specific character of evolutionary theory, if we consider that its explanatory strategies are always related to some use of the selectionist explanation, is that it brings together some formal models, written in mathematical language and in the modality of pure necessity, and some historical narratives, which allow scientists to instantiate the modes of natural selection in actual cases. Paleontology as well as the population genetics of given groups, species or clades, is essentially committed to a double-faceted scientific conceptuality, both historical and nomothetical.

The methodological aspect of this status is that no empirical inquiry in evolution can be made without comparative data, or some incorrect ahistorical use of natural selection mechanisms, such as the one exemplified previously, is always possible. And often the divergence of results between inquiries has to be traced back to differences between the sets of comparative data used by the researchers (Sober and Orzack 1994, 2001; Griffiths and Sterelny 1999 pp 240–250).

This pervasive character of historical narratives in neo-Darwinism accounts for the historical meaning of all biological terms throughout evolutionary theory. Taxonomies are obviously and easily reinterpreted by history, and the concept of homology that helps systematists to build their classification immediately acquires the historical status of “common descent’s sign” like bird’s wings and bat’s wings. Homoplasy, as the other kind of similarity across species, seems at first glance less historical: similar selective pressures gave rise to similar devices, as adaptations to such contexts. But is the concept of adaptation lacking any historical dimension?

The question of whether adaptation is a historical concept is widely debated. Even if it is admitted that “adaptation ascription are causal-historical statements” (Brandon 1996), since to say that a trait is an adaptation is to say that it has been somehow selected for some of the advantages it gave to its bearer via differential

---

12 On phylogenetic inertia, see Sober and Orzack (1994, 2001), Reeve and Sherman (2001), and below 2.2.2.

13 A case for historical narratives in evolutionary theory is made in Richards (1992); compare with the critique by Hull in the same volume. Gayon (1993) addressed the dual character of evolutionary theory.
reproduction, it remains to be decided whether this provides the whole meaning of the concept—given that, in fact, biologists often do not appeal to historical inquiries in order to describe adaptations but just forge optimality models with current data. Reeve and Sherman (1993, 2001) advanced a powerful “current concept” of adaptation, opposed to the historical concept sustained by the majority of philosophers of biology in accordance with Brandon (Sober 1984; Ruse 1986; Griffiths 1996). They propose that: “an adaptation is a phenotypic variant that results in the highest fitness among a specified set of variants in a given environment” (1993 p 9). Having distinguished two goals of evolutionary research, the first being the explanation of maintenance of traits and composition of population, and the other the reconstitution of history of lineages, they argue that the former essentially needs the current concept, while the latter is much more related to the historical concept. Notwithstanding one’s conclusion, the fact that there is a historical component of adaptation ascriptions is important, since it allows biologists to distinguish between the origin of a trait as an adaptation and its current presence and maintenance.14 A trait that is adaptive might not have emerged as an adaptation or might have emerged as an adaptation for some other use: this is grasped by the concept of “exaptation” suggested by Gould and Vrba (1982), an example of this being the insect wings that probably emerged as thermoregulatory devices (Kingsolver and Koehl 1989). Exaptation has proved to be a useful concept for understanding a lot of features that appeared during the evolution of hominids (Tattersall 1998).

However, one should not conflate two meanings of the historical characterization of adaptation: there is a definitional meaning and an explanatory meaning. On the one hand, selection defines adaptation, since a trait being an adaptation means that it originated through natural selection. But on the other hand, we say that selection explains adaptation: is not this contradictory to this definitional meaning? Not at all, because by this we now mean that the explanation of a given adaptation may search for concrete selective pressures in the given environment, and this provides an agenda for experimental testing of hypotheses. An example of such experiments, and then of the explanatory use of the historical concept of adaptation, is the study by Antonovics of the differential sensibility of plants to gradients of metal in soils (Antonovics and Turner 1971).

14 Sober (1984), followed by some writers, calls it “adaptedness” in order to distinguish it from “adaptation”.
2.2.1.4 A formal characterization of selectionist explanation and the issue of units of selection

The precise characterization of adaptation has been the focus of a large controversy in biology, and the philosophy of biology, about what exactly is likely to be adapted. Wynne-Edwards (1962) claimed that groups and populations were sometimes adapted. This meant that natural selection worked in favor of groups or populations, which sounds contradictory to the fact that individuals are the entities subsisting and spreading through selection. Williams (1966) gave a forceful defense of selection and hence adaptation as bearing exclusively on individuals. And, since the determinants of hereditary variations that are selected are genes, he concluded that selection acts primarily at the level of the genes. It is to be remarked, however, that he did not refute the logical possibility of adapted groups and then group selection, but proved that the alleged cases of group selection were explainable by natural selection at the level of genes, which is somehow more theoretically parsimonious.15 Along those lines, Dawkins (1976, 1982) elaborated his view of genic selectionism (or the “gene’s eye view” of evolution), trying to account for all manifestations of selection.16 One must here distinguish genic selectionism, which is an assertion about processes of selection and selectionist explanation, from genic determinism, which means that all phenotypic traits are wholly caused by genes with no impinging of environment or learning. One could perfectly subscribe to genic selectionism without genic determinism, as did Dawkins himself, and Dennett (1995) or Rosenberg (1985). Some practitioners or propagandists of sociobiology did make this confusion, which bore some hazardous moral and political consequences that in turn proved quite damaging for serious research conducted in those fields. But the genic selectionist’s concept of gene only requires that the presence of a gene in a genotype in an environment makes a difference relative to the lack of this gene in the same context; the weaker requisite is that genes be “difference makers” (Sterelny and Kitcher 1988), and here there is no commitment to any assumption about what genes determine and through which channels. This idea is the meaning of the locution “gene for” which unhappily has been read in a deterministic sense. Genic selectionism does not prevent an environment from being as much a determinant as genes in the success of a trait (Gray 2001).

15 In fact, Williams (1992) elaborated and defended the concept of clade selection, added to gene selection.
16 On the differences between Williams’ conception and Dawkins’ gene’s eye view, see Kenneth Waters (1991).
One of Dawkins’ major arguments was the concept of inclusive fitness developed by Hamilton through his research on kin selection. Here, turning to the level of genes within selective explanation appeared fruitful in studying such features as cooperation or altruism that sounded at first glance contradictory to natural selection as enhancing the individual’s fitness. The question, then, was to determine a level at which natural selection could explain the fact of altruism such as sterility of males in some hymenoptera species. Altruism has been selected because, although it decreases the fitness of the altruistic individual, it increases the representation of its genes in the next generation, provided that the individual is closely enough related genetically to individuals benefiting from this altruism. This is the case in insect societies that are essentially kin societies.

Genic selectionism has been challenged in several ways by Gould, Lewontin, Sober, and Brandon. One main argument is that selection acts only on phenotypes hence is blind to genotypes. Therefore, the level of genes is not relevant for understanding selection. Many genotypes, hence many genes, are identical selectively provided that they are “genes for” the same phenotypic trait. Using a notion elaborated by Reichenbach and Salmon in a philosophical debate about probabilities, the argument states that phenotypic interactions screen off the efficiency of genotypes and their relationships with environment. It does not deny that together with environments genotypes cause the phenotypes, but rather that this kind of causation does not explain the outcome of selection since it is necessary and sufficient for this purpose to consider the effects of the interaction of phenotype with its biotic and abiotic environments.17

The other line of defense, stated by Lewontin and Sober (1981), is the context-sensitivity principle, which claims that, since the phenotypic effects of a gene depend on the environmental and genetical context of its expression, a single allele cannot be the bearer of the selective causal process. The authors’ example is then the case of heterozygote superiority, since in this case the diploid genotype (e.g., AA, or Aa or aa) and not the single allele (e.g., A or a), is the genuine entity supporting the selection process. One can surely mathematically describe what happens to the single allele, but this gene’s eye view account is not causally explanatory.

Although biological evolution has been theoretically defined as a change in gene frequencies, the fact that the general model of the process of selection is not committed to any determination of the entities undergoing natural selection

17 A screens off B as a cause of C iff Pr (C/A&B) = Pr (C/A) ≠ Pr (C/B). Sober (1992) contested that “screening off” can yield a rebuttal of genic selectionism, since the argument is open to an almost infinite regress within which sometimes the most important explanatory cause is not the one which screens off all others.
implies that they are not necessarily “genes.” Hull gave a formulation in terms of replicators (hereditarily reproducing entities) and interactors (entities whose causal relationships affected the hereditary success of the replicators they are associated with). In most classical cases of natural selection, replicators are genes and interactors organisms. The gene’s eye view, then, says that since genes are the replicators, they are the units of selection. But of course this formal definition could be applied to other cases, in which replicators could be species or clades, or interactors could be genes themselves or groups. A more fine-grained approach of the various processes of selection is, thus, allowed by this formal characterization. For example, in the case of meiotic drive or segregation distorters—cases important to Dawkins’ argument—genes are themselves the interactors. But the “replication” idea faces some specific problems since it mixes the idea of reproducing, and the idea of copying (Godfrey-Smith 2000a). Only genes replicate since organisms reproduce but they do not copy themselves; yet, unlike organisms, genes are not alone in the replication process, they are involved in a whole machinery (ribosomes, enzymes, proteins, etc.). However, this machinery allows the differential expression of genes in the genome through regulation of their transcription into mRNA and proteins. This transcription might be thought of as a copying of the gene, but it is not really a reproduction, since it is a process distinct from the replication of the cell in mitosis or inheritance through meiosis. Moreover, this copying process is submitted to the regulation of gene expression, contrary to the replication in mitosis. The general conclusion is that in no case are “copy” and “reproduce” synonymic or correlative notions, which weakens the very notion of “replication.”

One point of interest of such a formulation is nevertheless that it can handle selection even outside biology, for instance, when we talk about cultural entities. The theoretical problem that faces this vision is to define a form of heredity in order to pick out the replicators. Hull thought that his formal characterization of selection was quite a logical one and embraced all possible cases. Giving the counterfactual example of the “protein world” that has no replication of entities, Godfrey-Smith (2000a) showed that the interactor–replicator couple is not necessary to the selection process as such; however, in our world, almost any selective process does rely on these elements.

---

18 Dawkins coined “replicators” but opposed it to “vehicles”; this word is too bound to the intuitive organism-genes difference.
19 For a general abstract account of theories using the concept of selection, such as immunology or evolutionary theory (“selection type theories”), see Darden and Cain (1988). For a theory of selection forged to address both evolution, immunology, and operant behavior, see Hull et al. (2001).
Easily expressible in this context, another argument against genic selectionism rests on the “parity thesis” (Sterelny, Griffiths), stating that all elements of the replication process are on a par (Griffiths and Gray 1994; Griffiths and Knight 1998; Oyama 2000), since environments as well as cytoplasmic elements or learned traits are “difference-makers” in the phenotypic outcome, exactly like genes. This thesis challenges both genic selectionism and genic determinism. Moreover, proponents of this theoretical alternative sometimes called Developmental System Theory contend that there exist kinds of heredity other than genetic inheritance, for example, nest styles, bird songs, or methylation patterns (Gray 2001; Jablonka 2001; Neumann-Held 2001). This challenge to gene’s eye view, consisting in a multiplication of the replicators, is more radical than the other critiques because the whole concept of the selective process has to be transformed. However, while accepting several kinds of replicators, Maynard-Smith and Szathmary (1995) trace a line between “limited” and “unlimited” inheritance, the latter allowing a quite infinite range of creation and transmission of elements. Only genes—and language—provide such an inheritance, which accounts for the extreme diversity and creativity of biological and cultural evolution.

The thesis of genic selectionism has been undoubtedly stimulating in compelling people to clarify their concepts and presuppositions. In fact, reacting to Dawkins’ extreme positions, some biologists did conceive of cases and mechanisms of group selection that could escape Williams’ critique. Nunney (1999) tried to define lineage selection, an idea that was suggested by Hull (1980), whereas Gould invoked a species selection (for properties, such as size, sex, etc.) that Williams (1992) refuted, although admitting and defending a clade selection above the level of genic selection. D. S. Wilson and Sober provided theoretical grounds for the use of group selection (Sober 1988a; Wilson 1992; Sober and Wilson 1994, 1998), and, especially in Unto others, designed a pathway from evolutionary altruism to psychological altruism. Their argument first relies on

---

20 The gene’s eye view defense has been the notion of information, in order to qualify the specificity of the genes’ role against other factors. Information should be semantically defined. It is not a simple usual correlation (fire–smoke) that is always reversible. Maynard-Smith (2000) elaborated this option, but difficulties raised by Godfrey-Smith (2000b) are numerous and go against a univocal notion of biological information.

21 Clear formulations of the DST program are given by Gray (2001) and Oyama (Oyama, Gray, Griffith 2001). On the developmentalist challenge and its integration into evolutionary theories, see later.

22 Gray (2001) provides a critique of the distinction.

23 Williams’ argument is that only a set of gene pools can behave in the same way as a gene pool when it comes to natural selection. Hence, there is only clade selection above the level of gene selection.
distinguishing and comparing within-group and between-group selection processes. Wilson and Sober’s argument invokes a “common fate” (Sober 1988a) of individuals in a group selection process, which implies that the selection process is compelled to act on all those individuals as a whole; then, secondary selective processes maintain this common fate, and selection can act at the level of the group. One consequence is that even kin selection appears as a form of group selection, rather than being genic selection’s underpinning of an apparently altruistic phenomenon. Group selection, however, is not exclusive of genic selectionism, since its point is that groups are vehicles; it is yet another question to decide whether or not genes are the only replicators involved.

Concerning genic selectionism, two strong positions are opposed nowadays among philosophers of biology. The first one, formulated by Brandon (1988) is pluralism: it states that there are several levels of selection and several units of selection. It is then an empirical question to know in any given case which are the actual forms of natural selection, but most empirical evidence is in favor of selection above the level of organisms in some cases, added to selection at the level of genes. The opposing position is defended by Sterelny and Kitcher (1988), who claim that there is always a genic selectionism which operates together with any kind of selection, even if we cannot have empirical access to this level, and even if it is pragmatically more interesting for biologists to recognize supraorganismic selection processes and treat them as such. The genic level is always the “maximally informative” one.24

Those philosophical considerations do not, in fact, impinge on biological investigations. It seems that biologists are in practice mostly pluralist on this issue (for example, Williams 1992), but it is not clear whether the decision between the two contrasting positions could be settled by the results of empirical inquiry. Some biologists, in fact, ignore those considerations and take for granted, since it is required by their practice, that there are several levels of selection (Keller and Reeve 1999) that have to be studied for themselves. But the recognition of the plurality of levels—notwithstanding the question of its ultimate theoretical reducibility to a genic one—gave rise to the important biological issue of their articulation. Michod (1999) elaborated the schema of a Darwinian dynamics, which accounts for the progressive emergence of new kinds of units of fitness: macromolecules, genes, cells, organisms, etc. The process relies widely on trade-offs between decrease in fitness in lower levels (for example, association of

---

24 Lewontin and Godfrey-Smith (1993) showed that even if an allelic descriptive and predictive model is always possible in cases classically opposed to genic selectionism, provided that one in some case enriches the model with conditional probabilities of alleles, however, those formal questions of adequate models do not decide the point of what is the causally relevant level.
individuals creating a common interest, which hurts the interest of the individual) and increase in fitness at the higher level (for example, the level of the association itself), and this trade-off is exemplarily a case of multilevel selection. The recurrent problem is then to find models that show how, in each case, the prime for defection (e.g., breaking the association), which is available each time there is a “common good” (Leigh 1999), can be overcome through this multilevel selection.

Although developed at a rather conceptual level, and mostly by philosophers, such controversies bear important consequences for the general meaning of the theory of evolution. What is at stake with altruism is the possibility of extending selectionist explanations in order to understand phenomena in the human domain. If altruism is explainable either by kin selection theory, or by Trivers’ reciprocal altruism (1971), which holds for populations of nonrelated organisms and is now derived from Game theory, particularly from the results of the study of the Prisoners Dilemma by Axelrod, then the issue of levels and units of selection is at the same time the issue of the foundation of an evolutionary approach, not only of the emergence of man and human societies but also of the current human psyche and societies through a selectionist framework, a research program now called “evolutionary psychology” (see below 3.2). Of course, altruism as studied by biologists is not what vernacular language calls altruism. For example, some very “egoistic” fellow (in ordinary language) would be biologically altruistic if he also wanted to leave no offspring; in contrast, a mother who sacrificed an entire life to her children, even if the perfect model of “altruism”, would from a biological point of view be typically selfish since she is entirely devoted to entities, which share 50% of her genes.25 So, no matter what the conclusion of the units of selection debate, all the lessons that might be taken from evolutionary biology into psychology have to be checked regarding whether they use vernacular or technical concepts and whether they do or do not carry illegitimate confusions between those two meanings.

2.2.2 Limits of selectionist explanation

2.2.2.1 The debate on adaptationism

Even if selectionist explanation is capable of rendering intelligible many nonbiological facts, no matter how far this capacity will be proved to extend, there

---

remains the preliminary question of its limits within the field of evolutionary biology. Darwin said that the Tree of Life (first principle of Darwinism) was partly explained by natural selection (second principle), but that there were other mechanisms at work in its production. So, I now turn to the actual limits of selectionist explanation in explaining both the form of the Tree of Life and the peculiar features of organisms.

The question of the limits and conditions of selectionist explanation has been approached in what has been called the controversy about adaptationism. In a very influential paper, Gould and Lewontin described and criticized a too pervasive method in evolutionary biology, which they called “adaptationist program.” In short, adaptationism means to think that the majority of the most important features of the living realm are explainable by natural selection (Sober 1994a).

There have been many attempts to clarify this adaptationism (Sober 1994a; Amundson 2001; Godfrey-Smith 2001b; Lewens forthcoming) 25 years of spandrels. In the cited title, this program, in summary, consists of atomizing an organism into discrete traits, and then building a selective history which establishes how each trait appeared as an adaptation to solve a peculiar problem. The authors contend both that we cannot atomize an organism in any way we want and that each trait allows the reconstruction of a selective history which should be testable. Too often, biologists, be they ecologists, paleontologists, or ethologists, create “just so stories,” e.g., stories that invent a plausible scenario of the resolution of a supposed antique problem—the trouble being that there is no way to prove that such a problem existed.

I have no interest here in deciding the fate of adaptationism. In fact, the most salient consequence of the spandrels paper is the necessity of clarifying the implicit assumptions in the research on adaptations, leading to a real formulation of an adaptationist program, and forcing scientists to take sides on the question.

26 “I am convinced that natural selection has been the main but not exclusive means of modification.” (1859 p 6). First of all, sexual selection, that is here left aside. For its complex relationship with natural selection see Mayr (1965a) and then the current research on the evolution of sex (Williams 1975; Maynard-Smith 1978).

primarily through the questions of its limits within biology. And here the major
concept pointed out by Gould and Lewontin is “constraint.” By this word, people
mean various things and state of affairs, so some clarifications are needed.28
Constraints can be physical, such as the size of the genome, which entails some
impossibilities for rapid metabolism within a cell in salamanders (Wake 1991); or,
more obviously, an elephant cannot have thin feet. They can be of genetical order,
for instance, when two genes are too close to be separated by crossing over during
meiosis. They can be phylogenetic, meaning that selection acts on entities that
come from a determinate history and have then inherited features difficult to
change. For example, selection cannot adapt a respiratory system of vertebrates
by creating a perfect respiratory device but has to modify the preexisting devices
in fishes. A constraint is recognized by comparison across several species or
clades: the fact that giraffes, like all mammals, have seven neck vertebrae like
mice, indicates that the number of such vertebrae is a constraint, since we would
expect number of vertebrae to be more proportioned to size (therefore more
adapted). Moreover, phenotypes undergo genetic constraints, since there are
epistasies and pleiotropies which entail that a trait will, in any case, be accom-
panied by another trait which has no adaptive relationship to it. Or constraints
can be “developmental”—this word needs some further comment. Those mean-
ings, unfortunately, are not easy to distinguish in fact. But let us keep in mind that
this issue of the limits to the power of natural selection (of a given trait) facing
constraints is tightly bound to the other issue raised by Gould and Lewontin,
namely the impossibility of atomizing living beings into discrete traits. The set of
constraints in the end gives the conditions for a kind of form untouched by
selection but always slightly altered and reshaped by it, which after the German
morphologists Gould and Lewontin named Bauplan.

However, the emphasis on constraint should be best understood with refer-
ce to the recent evolutionary theory of development.29 Selection acts on
variants, but not all variants are able to develop from a given gene pool. The
evolutionary theory of development unveils the constraints on the rise of those
variants upon which selection is about to act. For example, Wake (1991 pp 547–
549) showed that in all species of plethodontidae, the feet have got four toes
instead of five in the ancestor from which they derived by miniaturization. This
happened in unrelated lineages, as an alternative state of developmental mechan-
isms sharply distinguished with the five-toes producing state. Adaptive processes

text.
select for size, and developmental constraints switch from five to four toes, independently of the lineage.

This example has nonetheless been challenged by Reeve and Sherman (1993) in one of the most convincing defenses of the adaptationist program. Their argument is rather simple: one can always appeal to selection even in Wake’s case, since it is possible that there is a selection at an embryonic stage that eliminates variants that have more than four toes. So, the case for developmental constraints is not so easy to defend in the face of elaborated and differentiated concepts of natural selection.

Wimsatt elaborated the helpful concept of generative entrenchment, meaning that, no matter whether or not selection acts on some feature, the fact that it has been built into the developmental program of a species at a rather early stage implies that it is easier, less costly, and more probable for selection to modify traits that appear later in development. Since to modify a very entrenched trait entails modification of numerous connected traits that are built on it, this modification is very likely to be nonadaptive, hence disregarded by selection. The more relative to the early formation of the organizational plan of a species a trait is, the more entrenched it is, so the less probable it is that selection will act upon it and modify it, hence it can be considered as a constraint for selection.

Clarification of this debate has been provided by Amundson (1994, 2001), by arguing that in the end, developmentalists and selectionists do not ask the same question. Selection is appealed to in order to explain why such and such variants arose and spread in the gene pool among a given set of variants; but developmentalists, on the other hand, try to answer the question of the nature of this set of variants: why are there these variants and no other variants, and to what extent is the emergence of some variants unlikely or impossible? This, in fact, is not exactly a constraint on selection because selection is an explanans to a different explanandum than the one developmentalists are interested in.

This recognition of pluralism within the various explanatory strategies in evolutionary biology is likely to eliminate the false problems created by the adaptationist debates and leaves philosophers and biologists with the task of formulating and evaluating what could count as an adaptationist research program. Following Godfrey-Smith (2001b) and Lewens (in press), it is useful to define two large categories of adaptationists, the empirical one, who makes assertions on the pattern of the Tree of Life and the actual mechanisms of evolution, and the methodological variety of adaptationism, who contends that

30 Wimsatt and Schank (1988).
31 Reservations are made about the generality of entrenchment by Raff (1996); a general critique of the concept is to be found in Sterelny (2000 p 377).
biologists have to suppose, first, the presence of adaptations, even if they recognize later that in fact the predefined adaptational optima are not reached and that constraints exist.

However, notwithstanding conclusions about the compared values of the many adaptationist programs or hypotheses, there is a larger fundamental issue to be addressed as a background to this question, namely the conditions under which we are likely to recognize the effects of selection and its place in relation to the other causes of evolution. I will first address the question of phylogenetic inertia related to selectionist explanation, and then I will turn to the question of the status of genetic drift.

### 2.2.2.2 Selection, drift, and phylogenetic inertia

Any model of real phenomena has to state a null hypothesis, namely, the description of a state in which there is nothing to explain, compared to which the actual state will have to be explained. Many radical changes in scientific thought, be they called “revolutions” in the Kuhnian sense, or more modestly “shifts,” consist in new definitions of the null hypothesis. For instance, Galilean physics began by conceiving of the rectilinear uniform motion as the “null hypothesis” (instead of rest), pointing out acceleration or trajectory changes as the right explanandum. And such a definition of the null hypothesis has been called “the principle of inertia.”

So, the words themselves suggest that phylogenetic inertia is the null hypothesis in evolutionary theory. In any population, traits have to be explained if they are not obviously the result of descent, e.g., if they are not homologous of traits in the ancestor species. Of course, the determination of the traits as homologous or not depends on the set of species that are to be compared. Thus, the preliminary definition of this set of compared species in order to account for traits in a given species is an absolute condition for applying the PNS. Wrongly determining homologous traits by inadequate specification of the initial set of related species to which the explanandum species has to be compared immediately entails false results (Sober and Orzack 2001). The “just so stories” stigmatized by Gould and Lewontin as unfalsifiable and abusive applications of the PNS often stem in the absence of data from such misunderstanding of the right null hypothesis.

However, methodologically, for a set of species, the relationship of homology and homoplasy implied by the statement of a null hypothesis is epistemologically related to the more fundamental principle of parsimony.\(^{32}\) It can easily be seen that

---

\(^{32}\) See this volume Chapter 5 by Folinsbee KE, Evans DC, Fröbisch J, Tsuji LA and Brooks DR. For a general philosophical account see Sober (1981, 1988b).
the more we judge there to be homologies, the less evolutionary lineages we have to draw on: this is a kind of parsimony, so Hennig’s auxiliary hypothesis can be called upon if one subscribes to epistemological parsimony. But the stronger, ontological claim of parsimony also supposes this way of defining the null hypothesis.

Phylogenetic inertia, however, is not incompatible with selection. We have to distinguish between the question of the origin of traits and their presence. When the traits exist by phylogenetic inheritance but decrease in fitness in the new environment and the new species, selection is likely to suppress them or render them vestigial. In the reptilian family, this was probably so in the case of the four legs when it came to the snakes. So origin and presence are two distinct topics. If the inherited traits are still present, one is allowed to postulate no negative selection, but a positive fitness value may promote stabilizing selection to keep them. So selection and inertia are not two competing hypotheses but are sometimes distinct explanans for distinct explananda, and sometimes complementary explanatory resources. In this regard, the idea of a null hypothesis in the question of the maintenance of traits has even been challenged (Reeve and Sherman 2001).

As early as the 1930s, Wright forcefully emphasized the evolutionary role of random processes (Wright 1932) such as genetic drift. The smaller a population is, the more powerful those kinds of processes are. This is a rather simple idea, since the same phenomenon is illustrated by the toss of a coin: a small sample is more likely to show a random bias (for example, seven heads versus three tails), than a large sample, which will show a half/half distribution of tails/heads, according to the law of large numbers in probability theory. So, in small populations, some genes, whose fitness is either equal or lower than other alleles, can go to fixation.

However, the concept of drift is not only a negative one, if it is connected with Wright’s other concept, the adaptive landscapes. The fact is that in a gene pool some combinations are local optima, and if a genotype is on the slope of this kind of local optimum, selection will lead it toward this peak. But there may be a fitness valley which separates it from another, higher, fitness peak so that its fitness will have to decrease in order to get it onto the other fitness peak. For this reason, only random drift, provided that population is small, can lead it through decreasing its fitness across the fitness valley toward another hill so that selection can, after that, lead it toward the global fitness peak. Then, through migration, the new genotype can spread. In this model, drift helps to increase fitness by moving genotypes to global fitness peaks. Drift is, then, together with natural

33 For the several interpretations of the adaptive landscapes see Gilchrist and Kingsolver (2001).
selection, the other process accounting for the evolution of species, modeled by the travel of genotypes across fitness valleys and hill climbing. Wright named this schema the “shifting balance theory,” and empirical evidence for its generality is sometimes given but is not generally persuasive.34

Of course, this goes against Fisher’s formulation that average fitness is always increasing; but the conditions of the two assumptions are not the same, since Fisher’s theorem speaks about large, theoretically infinite, populations. Hence, evaluating the conflict between Wright’s view of the role of random drift and Fisher’s claim of an overall selectionist view, according to which the fittest always invades the gene pool, entails a decision on whether large or small effective populations are mostly to be found in nature.

But random drift raises some epistemological questions (and a more metaphysical one that I will address in the next section). The issue is stating the difference between drift and selection: are they two competing hypotheses? A first model, explored in detail in Sober (1984), takes drift and selection as two kinds of forces acting on an equilibrium model formulated by the Hardy–Weinberg law (together with the forces of mutation and migration, which I do not consider here). If equilibrium is changed, then selection is acting; when the fitter allele is not fixed, then random drift must have perturbed selection. Outcomes are the result of the addition of selection and drift in an analogous manner to summation of forces in Newtonian mechanics. However, this model has been challenged in two papers by Walsh, Ariew, Lewens, and Matthen (2003, 2002). They argue that selection and drift are not equivalent forces since they are not as comparable as two directional forces in dynamics. They do not compete at the same level, because “natural selection” is not exactly a force like the sum of selection pressures, but a sampling effect, supervenient on the real selective forces (fight, predators, mate choice, foraging, reproduction...) in the same way that entropy supervenes on microphysical states of molecules. On the other hand, drift is another kind of sampling, in the manner of a sampling error (compared to the fitness coefficients). Since summation of forces presupposes that their effects are additive, therefore, are acting at the same level, one cannot logically treat the state of a gene pool as the composed effect of selection and drift; and, finally, to talk of forces proves in general to be misleading, even for selection.

The question is not an empirical one but concerns the logical types of those population-level theoretical entities or processes that are selection and drift. Epistemologically, this means that there may be such a gap between drift and selection that the model of composition of forces has to be replaced by a

34 Coyne et al. (1997).
thermodynamical model of macroscopic effects of the statistical accumulation of heterogeneous microphenomena. The analogy of selection is no longer gravity but entropy, and we know that entropy as a variable bears no causal effect. Whether or not Walsh, Ariew, Lewens and Matthen’s challenge is right, the point is that considering drift leads to no obvious model of selectionist explanation, since when one is about to derive empirical content from mathematical models of population genetics and the PNS, one has no sure principles for conferring an epistemological status to the process of selection. This does not affect our study of phylogenies or our making and evaluating models for its mechanisms, but the interpretations of those models, hence of the very nature of the mechanisms, are certainly at stake. If drift and selection are not to be compared as two different forces like electromagnetism and gravity in physics, Darwin’s statement about the composed nature of the processes of evolution, and the subsequent agenda of weighting the components, has to be qualified.

2.2.2.3 The scope of natural selection

Evolutionary theory puts together the Tree of Life claim, and the Selection principle; however, these two statements are not logically connected. We can imagine a possible world where there is selection and not one Tree of Life; Lamarckism gives a picture of the opposite possible world. The question then is the relationship between the two claims: to what extent is the Tree of Life accountable for by natural selection? This question was present but quite attenuated in Darwin since he thought of other mechanisms than selection (e.g., Lamarckian inheritance). However, it becomes urgent in the Modern Synthesis because it focuses on selectionist explanation in the forms and conditions outlined previously.

All the puzzles investigated in the preceding section concern the selectionist explanation in general, whether it is applied to speciation in a population and on a short time scale, or to what Mayr called “emergence of evolutionary novelties” (1959a) (such as the transition of the protostomes to deuterostomes). However, there is a difference between those two objects, and this raises another question about the scope of the selectionist explanations under consideration up to now. For instance, it is plausible that Wright’s SBT accounts for a lot of speciations on small time–space scales but that evaluating its validity on a wider scale may appeal to other criteria. Paleontologists distinguished after Goldschmidt (1940) micro- and macroevolution and wondered whether the same processes should be held responsible for the events in those two cases. Simpson (1944) argued that, even if macroevolution shows very different rhythms in different lineages, however, it
implies the same processes as microevolution. The main objection, from a paleontological perspective, namely the lacuna in the fossil records on a large timescale, could be explained by purely geological reasons with no need to postulate special processes to account for them. But Simpson felt compelled to isolate a “mega evolution”—e.g., emergence of new lineages—that cannot so easily be interpreted along the lines of microevolution. Of course, Eldredge and Gould were the most convincing proponents of the difference between macro and microevolution, with the paleontological theory of punctuated equilibria. This theory is, first of all, a reading of the fossil records which claims that discontinuity is not a result of geological lacunae (as Darwin tried to establish in the Chapter IX of the Origin) when they show no major transformation for a very long period of time followed by sudden change. Here, the process accounting for this record is interpreted as a dual one, composed of fine tuning adaptation, which is a kind of stasis, and then a quick general transformation of the body plan giving rise to a new phylum. If the first process is explainable by selectionist explanations, such as the one I have considered up to this point, the second stage needs at least a change in the conditions under which natural selection can operate—if we still assume that no other process is needed.

No doubt challenges to Darwinian gradualism were numerous before Eldredge and Gould: before the Synthesis there were saltationists like De Vries and afterward came the “hopeful monsters” proposed by the geneticist Goldschmidt (1940). As a result of this, Mayr (1965b) established that gradualism—meaning that no evolutionary change is due to a big mutation—is compatible with evolutionary novelties since any change (like exaptations of insect wings) or intensification (as in the evolution of eyes in some lineages) of function can account for many structural novelties. Punctuated equilibria is a really challenging theory because the difference in the form of the Tree of Life cries out for a difference in the nature or the conditions of processes. If we subscribe to the idea of Baupläne as an integrated set of constraints, as advanced by Gould and Lewontin (1978), then we might think that phases of stasis represent fine adaptive tuning of the existing Baupläne, whereas quick transformations represent the appearance of new Baupläne.

Nevertheless, this view rests on some orthodox considerations of selection: among the founders of the synthesis, Mayr (1965b) emphasized the stabilizing role of selection, which, given a particular environment, largely eliminates big mutations since, given the high degree of integration of most organisms they are probably deleterious and often likely to threaten functional integrity. Periods of stagnation are, therefore, to be expected, by the nature of selection. The crucial point, however, is the logical relation between large-scale and small-scale evolution. Founders of the synthesis, like Fisher and Wright, focused on
microevolution. However, some assumptions defining such evolution become false when we jump to macroevolution: environments are no longer stable, and they can change quickly and intensively; and phenotypic variation available is not stable either since a very different range of variation will be available if the time scale is larger.

This second parameter is connected to Gould’s other main concern, namely evolutionary theories of development, and the focus on heterochronies crucial to his *Ontogeny and phylogeny*. The question is: what are the constraints on the range of variation, and what constraints are about to change? Developmental constraints are likely to account for the restriction of available variation and then for the focusing of selection process upon fine adaptive tuning and finally for the puzzling outcome of stagnation in the evolutionary tree. If we want to understand the transformation phase, we have to turn to the modification of available variation and then to a possible change in constraints. To this extent, if a modification happens in developmental mechanisms, then we could expect an enlargement of phenotypic variation, a new field for selection and thus new evolutionary possibilities. This is because, if we consider that the features yielding this enlargement are deeply entrenched, we can understand that in this case selection will act upon many connected traits at many levels of the developmental process, and so a radical change of existing body plan is likely to result. This was Gould’s (1977) point, following De Beer (1958) concerning heterochronies: a change in the timing of development, involving many subsequent and connected transformations in the life cycle is more likely to transform the body plan of a species than is change in an adult trait. This sets the agenda for other kinds of evolutionary research, including not only the taxonomy of different mechanisms able to affect development and thus yield evolutionary novelties, but also an attempt of causal accounting for them (an agenda which is a part of the Evo-Devo program). The important discovery of *Hox* genes developed in Lewin’s studies on *bithorax* gene (1978; see Gehring (1998) for a historical account), which are homologous in arthropods and chordates, supports this thesis, since a slight replacement of such a developmental gene by the *Antennapedia* gene can give rise to a leg instead of an antenna in *Drosophila*. Setting aside the complexity of the cascades of interactions, the general idea is that great transformations of a Bauplan may be generated by slight modifications of some kinds of genes or of

35 Like in Gould (1977): heterochrony, paedomorphosis, neoteny are defined and exemplified.
36 Even if genes of this sort, such as Bithorax have been known since about 1915, a major stage in the emergence of Evo-Devo has been the molecular characterization of those genes in the 1980s, mostly by Gehring (see Gehring 1998). This revealed that homeobox genes are homologous across several phyla.
their expression channels (Arthur 1997) because the development and life cycles are affected at many levels. Whether this view will prove correct or will need a radical revision, such as DST claims, and no matter the range of biological cases to which they apply, its epistemological significance requires integrating developmental biology and evolutionary biology in order to assess the multiplicity of the processes needed to account for the varied features of the evolutionary tree.37

On large time scales, environments are very likely to change, not only due to the evolution of organisms and populations, but also because of general geological and meteorological shifts. This second dimension of mega evolution converges with the first one to present the philosopher (a) with an epistemological issue. It also inspired Gould in his stronger challenge to overall selectionism (b).

a. The epistemological issue is the following: when variation range and environment change, populations exhibit a response to selection constituted along parameters that were not previously relevant. It could be said, then, that populations and organisms are evolvable. But some features make them more evolvable than others. Hence, at this large evolutionary scale the question may no longer be the evolution of adaptations (with all the epistemological problems addressed previously concerning nature and limits of selectionist explanation) but rather the evolution of evolvability itself. Changing explananda, then, could shift interest toward other levels of selection than genes and individuals, for example, clades and populations, since some population-level traits, such as sex or polymorphism makes them obviously more evolvable (Gould, Williams, Sterelny). But it can also raise new questions, such as the evolutionary origins of those features of traits that make them easily evolvable: how, for instance, are we to explain the cohesion of genes in a chromosome (Keller 1999), modularity (Wagner 1995; Sterelny 2004) or redundancy? So shifting the scale in the Tree also shifts interest from epistemological and methodological issues proper to selection, drift and inertia, to a general concern with new objects, such as modularity.

b. In Wonderful Life, Gould tried to trace the philosophical conclusions of the recent analysis of the Burgess shale, particularly by Withington and Conway Morris. His verdict was that many phyla appeared with the Cambrian, of which only few survived; thereafter, very few new body plans and phyla were really “invented” through evolution. But this creativity in evolutionary novelty is somewhat puzzling and raises a concern for the new explananda stressed previously. With the famous metaphor of the film of life rerun, Gould suggested that the history of life was much too full of contingent events, such as the mass extinction that killed more than half of the Burgess phyla (plausibly after the fall of an asteroid, according to the Alvarez hypothesis). The punctuated equilibria claim was a weak challenge to an overall view

---

37 Gilbert et al. (1996).
of selectionism since it can be reinterpreted as the necessity of defining two regimes of selection, the second one including the aforementioned concepts and concerns stemming from developmental theory. This latter view presents a strong challenge, since selection, and the adaptive capacities of individuals and species, cannot prepare them to face mass extinctions due to excessively strong changes of environment. Hence, the ones that survived did not owe their survival to their higher fitness, and the explanatory and predictive power of natural selection is very limited at this level of the history of life. *Anomalocaris*, for instance, seemed quite well fitted to its marine environment and was undoubtedly a strongly performing predator, surely no less well adapted than *Pikaia*, which seems to belong to the chordate phylum; it nevertheless disappeared. Thus, major events are contingent with regard to the parameters ordinarily involved in natural selection. This “contingency thesis” heavily restrains the scope of natural selection.

The fate of this challenge rests on a lot of empirical elements that are not yet available. In particular, the diagnosis of the Burgess fauna is still debated since Conway-Morris himself revised his original judgment (1998) and estimated that many Burgess phyla are in fact ancestors of already known lineages. However, as Gould pointed out in his reply (Gould and Conway-Morris 1999), the point is not whether or not there are other mechanisms than natural selection, a conundrum that we are unable to solve, but whether there were many more new phyla in the Cambrian, a great part of which effectively disappeared. The contingency thesis relies on an affirmative answer to this question, which should be studied by paleontological and morphological means. So notwithstanding the strong challenge to selectionism, the important consequences for the interpretation of the history of life rely on empirical investigations. But the question is likely to be begged by methodological considerations involving disparity. If diversity means the variety of species, disparity means the heterogeneity of the body plans. Gould contends that whereas diversity may have increased, disparity decreased. But even if we knew what the Cambrian phyla were, this would not entail the ability to measure disparity (Sterelny 1995, 2000). Cladists mostly think that we can trace the genealogy of phyla, but not evaluate the distance or difference between two phyla, because the criteria are always instrumental. In this view, Gould’s thesis would not be testable. The basic question, beyond the measure of disparity, is the

38 For example, *Hallucigenia*, once viewed as a quite unique species in its phyla, if turned upside down could enter into the phylum of the echinoderms (Conway-Morris 1998).

39 An argument against the contingency thesis would be convergence, if similar features appear several times in different lineages, they are more likely to appear even if we change some initial conditions of evolution (Sterelny 1995; Conway-Morris and Gould 1999). But such an argument makes use of excessively undefined notions of necessity and identity.
counting of body plans, hence the definition of body plans. Failing any consensus about that, the contingency thesis, whether or not empirically adequate, is not likely to be tested.

From a distance, the current state of evolutionary theory may in general be characterized as facing two kinds of challenges, weak and strong. Weak challenges imply, if successful, a revision of some part of the theory in order to integrate new methods and concepts; strong challenges entail giving up some major credos of the Modern Synthesis. In the case of Gould’s punctuated equilibria and contingency thesis, those two challenges focus on the first Darwinian claim, the form of the Tree of Life. Here, the strong challenge would lead us to give up both gradualism and the hope of finding a general account of the history of life through one explanatory schema.

But the same situation arises in the case of the second Darwinian claim, concerning the process in evolution. Here, challenges are forged by developmentalists. The weak challenge proposed by Evo-Devo involves a rethinking of the conditions and mechanisms of selection when it comes to development and the origin of evolvability. The strong challenge is formulated by DST proponents and entails giving up the concept of gene or its main role in inheritance and selection.40

2.2.2.4 Preliminary assumptions concerning the view of selection

The controversies addressed here over the limits and scope of natural selection, although not devoid of empirical content, are largely dependent on the conception that the authors have of the nature of selection. So far, I have left aside the most general alternative regarding this conception, an alternative which provides both a negative view and a positive view of selection. In the former option, selection merely selects, hence it just sorts high-fitness traits against low-fitness ones; in the latter option, selection is by itself creative. Mayr (1965b) claims selection is not a “purely negative force” since it gradually improves existing traits. Among biologists, this positive view is widely held: Dobzhansky, Simpson, and Gould shared Mayr’s view.

The general question underlying this split is: what does selection actually explain? It does not explain why this individual has this trait (this is due to

40 Kitcher (in Gray et al. 2001) and Sterelny (2000) are moderate appraisals of the extent to which the developmentalist challenge demands revision of the evolutionary theory. Kitcher (2004) is also sympathetic with Gould’s weak challenge but defends neo-Darwinism against Gould’s strong challenge.
developmental effects); as a population-level explanation, it explains precisely why this trait, once it had arisen, pervaded and persisted in a population. Thus, Sober (1984) subscribes to the negative view that selection is a population-level explanation, as we noted, so that the question “why is trait A in individual B?” does not belong to what it explains.\(^{41}\)

Neander (1995) challenged this view in a paper expressing the epistemological substance of Mayr's intuition. Apart from the two questions that I have distinguished (the “developmental question” and the “persistence question”), there is the “creative question,” which is: why did the genetic and developmental devices underpinning a given trait arise in a population? Contrary to the positive view, Neander contends that natural selection contributes an answer to this creative question. Perhaps speaking of creation is misleading because of the connotations of the word, evoking an instantaneous happening. I give here a slightly modified argument. In fact, even if the genotype conditioning the new trait is not created by selection, selection does increase the probability of the several genes composing this genotype, given that some of its genes have already arisen. The point is that, if a high fitness trait has a genotype \(G_1 \ldots G_9\) (measured in a model of the fitness of possible genotypes), and if \(G_1\) spreads into a population, then without hypothesis of selection at all, \(G_2 \ldots G_9\) are not more probable (than other alleles) than before; but under the hypothesis of selection, once \(G_1\) is there, the probabilities of \(G_2, \ldots G_9\) being fixed are significantly raised since they are part of the higher fitness genotype \(G_1 \ldots G_9\) and are more likely to appear than the less fit \(G_1G_2G_3 \ldots G_9\). So selection has \textit{causally contributed},\(^ {42}\) not only to the spreading of the genes \(G_1, G_2, \ldots G_9\) in the population, but also to the emergence of the integrated genotype \(G_1 \ldots G_9\), namely, the novel trait we are considering.

This defense of the positive view of selection can be extended. Natural selection is a three-stage process, variation, differential reproduction as effect of the variations, and change in gene frequency. But the two first stages are not easy to distinguish, since, although variation is conceived as resulting from mutation and mostly recombination, selection may affect the regime of variations, and therefore, controls the very parameters of its own exercise. It has been shown in some bacteria exposed to stress that selection can enforce the mutation

\(^{41}\) Of course, the issue of explanation is correlated to the question of causation: “what does selection actually cause?” More strongly committed than Sober to the Negative view, Walsh (2000) claims that natural selection is not even a cause of adaptations, since strictly speaking their causes have to be found in the developmental life cycles of the individual and forces acting upon it. Selection is merely a sorting process that presupposes such real causes.

\(^{42}\) At least in the probabilistic sense of causation, meaning that a cause increases the probability of its effect.
rate, providing an advantage in the range of available selective responses to environmental shifts.\textsuperscript{43} The notion that mutation rate is somehow controlled by selection, whereas mutations are the material upon which selection acts, demonstrates a kind of reflective impinging of natural selection on its own parameters. Such a reflexive structure of selection allows one to say that the traits selected are themselves dependent on the form of selection pressures, hence that they are somehow shaped, not only sorted, by selection. In the present case, even if the content of mutation is not given by selection and is prior to selection, any individual mutation is still counterfactually dependent on selection since the probability of its occurrence is directly dependent on the mutation rate. In the same spirit, Mayr (1970) suggested years ago that competition itself is under the control of selection (since too much competition could render selection impossible). To this extent, it seems difficult to separate positive causes of the new individual phenotypes emerging (“shaping”) from negative causes affecting their spreading or extinction (“sorting”). So, provided that the causal and explanatory regimes of natural selection are conceived as different from the explanatory regime at the individual developmental level, the positive view of selection is likely to be adopted.

2.2.3 Metaphysical issues about natural selection: some problems with realism

At many times, the issues exposed here involved a general metaphysical question, which is the problem of realism. Under this name philosophers of science try to understand the status of theoretical entities, such as electrons, oxygen, energy, gene, etc.—entities, which are often nonobservable. Roughly speaking, some defend realism, which means that those entities, and the process that involve them, are real things, whereas others are instrumentalists or pragmatists, which means that those terms gain there meaning only in the context of the scientific inquiry, mostly to allow predictions and other tests.

The fact that natural selection itself could not achieve the status of general physical laws alerts us that the problem of realism could be different in biology and physics. Rosenberg (1994) convincingly defended the thesis that evolutionary biology, as opposed to physics, must be conceived of instrumentally. One of his chief arguments is the supervenience of every concept bound to natural selection. Since natural selection selects for effects notwithstanding their causes (e.g., it

\textsuperscript{43} Andersson, Schlechta, Roth (1998); Mc Phee, Ambrose (1996); Taddei, Matic, Radman (1995).
selects for function no matter the physical structure realizing it), according to a natural-selectionist’s point of view different real processes and entities are treated as the same thing, which implies that this is an instrumental perspective since it abstracts from the differences between those infinitely varied real processes. “Instrumental” here means that natural selection is a concept so coarse-grained that it misses the fine-grained distinctions between real processes, albeit still useful for us in making the depictions we are interested in, since the fine-grained knowledge of all those processes and entities is out of our grasp.

The case for instrumentalism arises also in the context of another issue addressed previously, namely the units of selection controversy. Arguing against genic selectionists, some authors (Sober, Brandon, and Gould) accept that processes, even of group selection, can be described at a genic level since in the end evolution is change in gene frequencies—while contesting that this is the description of what actually happens. Thus, they suppose that the selection process is real and not dependent on our cognitive interests, and in this case the question is to specify the exact level of this process. Realism makes the controversy over units of selection more pressing. On the other hand, pure instrumentalism would dissolve it into the methodological question of the best mathematical model for a given process.

Realism in this context means that there is one real process of selection, and we have to decide what identifies this real process. Realism, of course, does not preclude any option regarding the debate itself: one can be a realist genic selectionist, allowing group selection to be an interesting description of phenomena, but one that does not identify the real process. But due to the structure of natural selection, it is not certain whether this sharp distinction between a real process and a convenient description holds. Following a suggestion by Kenneth Waters (1991), I will stress some consequences of the context-sensitivity principle.

Single genes, considered at a population level, are selected for or against with regard to their context, which means the environment of their phenotypic effects and principally their genetic environment (Mayr 1965b). Their selective advantages are context-sensitive. But the argument applies, finally, to any presumed unit of selection: its fitness depends on the whole interaction. Even Dawkins (1982) uses it in order to reject the claim that nucleotides in the end could be the real units of selection (their context is the entire allele).

But given that we are considering a population of such entities, contexts (environments for organisms, genetic environments for genes, etc.) are not always likely to be homogenous. Even in the classical case of Kettlewell’s industrial

---

44 This is still the line of Sober’s reply (1990) to Sterelny and Kitcher (1988).
melanism, there are places in the woods where trees are mostly white, others
where trees are mostly black, and in each of those contexts fitness values of black
moths and white moths differ. Then the fitness value of the entity is obtained
through averaging the various values across the varied homogenous contexts.
Now, if we are tough realists and claim that only causal interactions in nature,
such as the selection or deselection of an entity within its given context, are real,
those averaged values will be mere convenient placeholders for the real processes.
But in this case—and this is the most important point—the argument holds
equally against genic selectionism and against organism-level selectionism, since
as averaged, all fitness values are such placeholders. So if we do not want to
collapse into total instrumentalism, we have to accept that selection processes
with averaged fitnesses at many levels are different ways for us to describe the
same real process and the unique way to get real information about it. But at all
those levels, different forms of information are complementary.45 Picking out the
supposed real processes, each in its single context, is not exactly the job of the
evolutionary theorist, but he or she has many ways to describe the same multi-
plicity of processes. If there is in fact one real process, it nonetheless must be
addressed at several descriptive levels. Kenneth Waters calls this a “tempered
realism” because it tempers the sharp distinction between real theoretical entities
involved in single-context selective processes and instrumental concepts.

This case for pluralism has to be distinguished from the multilevel selection
advocated by Brandon or (in another way) by Sober and Wilson and from the
Sterelny–Kitcher theory. In Brandon’s pluralism, there are many possible forms of
selection, but there might be cases in which selection plays only at one level.
Pluralism is then compatible with tough realism. In contrast, pluralism sustained
by tempered realism contends that there are always several levels of description
for a same process—context-sensitivity implying that there is no way of discrimi-
nating what is the “real” process from descriptive reconstitutions of processes.
This last assertion contrasts with Sterelny and Kitcher pluralism since these
authors claim that the genic level description is “maximally informative,”
and—unlike other types of selection—is in all cases available. Tempered realist
pluralism is not committed to such a genic privilege.

The controversy about units of selection was not expected to be solved by
those considerations, which had two interests: to exemplify the fact that episte-
mological essential debates in evolutionary biology bear important metaphysical
consequences and to illustrate the requisite of forging a definition of realism
proper to evolutionary theory when we are about to discuss those metaphysical

45 According to Kenneth Waters (1991) each description is dependent on the prior decision on how
to discriminate between environment and selected units.
matters. In this perspective, the idea of tempered realism should contain lessons for other entangled debates about the metaphysical and epistemological sides of evolutionary biology. Due in particular to the epistemological status of natural selection, no general assertion from philosophy of science can decide the issue of realism within it and even enunciate what it would mean to be a “realist” in this context.

2.3 An evolutionary framework for philosophical issues?

A philosophical focus on evolutionary theory cannot ignore the huge consequences that Darwinism had for traditional philosophical problems, ranging from theological and moral matters to psychology. Since this field is as wide as philosophy itself, I will not even try a small survey but will instead indicate two or three perspectives on the fecundity and the difficulties involved in the use of evolutionary considerations in philosophical debates on human nature. Evolutionary theory pervades the whole theoretical discourse on man: from philosophy of mind and language to morals through epistemology.46 Two general motives can be distinguished: applying the power of selectionist schemes to problems outside biology; integrating traditional problems of meaning and culture within a large evolutionary framework, which permits asking anew the question of the origins of some devices (of ethics, of language, etc.), which has been avoided because of the untestability of hypotheses. In a strict sense, only the second strategy can be called a naturalistic evolutionary framework; however, the first one is almost always thought of within a naturalistic strategy.47 Here I will not consider issues in evolutionary ethics (see Chapter 25, Volume 3).

2.3.1 Selectionist models of culture and science

As we have seen, there exists a general formula of selection that does not specify the replicators and interactors. Thus, provided that we can establish heritability for some cultural or moral entity, a selectionist schema could render its birth and fate intelligible. Culture, and science or epistemology, are the most important

---

46 The main general philosophical account of the consequences of evolutionary theory outside biology is Dennett (1995).

47 “Naturalism” qualifies here any research program that formulates its questions and constitutes its method along the sole lines of natural science methodology, and that does not accept any entity originating and subsisting for itself above the nature studied by those sciences.
fields of explananda for those theories. It was noticed long ago that one human characteristic, whether unique or just developed to a unique extreme, is culture; and the fact is that besides genetic inheritance there is cultural heredity: individuals learn, and they can transmit what they have learnt, which appears somehow replicated. So, given that the transmitted items are likely to be slightly modified each time they are reproduced, a selectionist evolutionary account of culture can be worked out. It faces at least two big problems: the first one is the relationship between genetic and cultural inheritance. The fact is that an entity may have great “cultural” reproductive success while its bearer leaves no offspring. So there is no necessary genetic basis for cultural traits, although if a cultural trait enhances the fitness of its bearer—think of any medical device that fights illness—then this enhances its own reproductive success. The other problem is the definition of this “success,” given that cultural transmission is not only vertical as in heredity, but also horizontal (e.g., toward non off spring), and this dimension is at least as important as the vertical one for the spreading of a trait.

Boyd and Richerson (1985) formulated a powerful set of models for cultural evolution. They did not make general assumptions concerning the genetic bases of cultural traits, but clearly cultural evolution, with the vertical and horizontal dimensions of transmission, implicates rather complicated relationships with genetic evolution. One of the main results of the models is that cultural evolution is far faster than genetic evolution. This strengthens, justifies, and explains our intuition that since long ago in human history, biological evolution has been negligible compared to cultural changes.

There is a large methodological issue here: is this selectionist theory of culture analogous to natural selection in biology, or is it a part of this theory? Boyd and Richerson are quite neutral; Dawkins’s “memetic” sounds like an analogon of evolutionary theory. Other authors (Lumsden and Wilson 1981) are more committed to a “continuous” view, which includes a criticizable assertion on the genetic basis of cultural traits.48

The part of culture which is technique is susceptible to a similar evolutionary approach, and it may be easier since the problem of adaptation has many parallels with technical evolution (Basalla 1988). If we consider technique (for example, photographic cameras since their origin), it is even possible to draw an evolutionary tree similar to a branch of the Tree of Life, with extinctions, radiations, privileged lines of evolution, and so forth. And concerning mechanism, we have variation provided by changing fabrication technologies, and we have a kind of

---

48 A severe critique is to be found in Maynard-Smith (1996). A general evaluation of the possibilities offered to selectionist theories of culture in a quite sympathetic perspective is given by Kitcher (2003b).
process of selection since the more robust or efficient objects are more likely to be copied. Hence, it should be possible to define a property analogous to fitness. It is plausible that this field of the study of culture will be most likely to receive satisfactory evolutionary treatment.

Science is a limited area of culture subject to a specific constraint (namely, its aim to represent the world in some way). The foundationalist program for the philosophy of science, in quest for a priori rules and guarantee for scientific inquiry, declined in the 1960s following extensive critiques and philosophers after Quine turned to a naturalized epistemology, e.g., an epistemology situated at the same empirical level as the sciences, not considering them from an a priori point of view. In so far as they use selection, which presupposes no trend among the selected entities toward any end, evolutionary models of science have the advantage that they do not presuppose any shared rationality or ideal from the scientists and even no special competence to recognize what is true. It is a fact that no definitive formulation of the goal of science, of what is “objective truth” and the criteria to recognize it, has ever been enunciated; thus, we cannot presuppose that all actors are oriented toward the same goal. A selectionist process is able to produce theories with the tightest match to the real world, whenever theories bear any consequences in practical life (those consequences will be the effects upon which selection acts) (Ruse 1986). At the price of giving up the idea that science aims at eternal and ideal truth (Giere 2001), evolutionary epistemology with a strong selectionist commitment, as originally formulated by Campbell and variously advocated by Giere or Hull (1988), gives a clear picture of the “process of science” that conciliates the lack of empirically attested “aim” with the cumulative improvement of the fit between theories, data and applications.

However, the nature of replicators and interactors in this process is a confused issue: contents (and how to define them), scientists, etc.—although the clearest account of this perspective is Hull’s. So, one has to prove that truth of theories defines a kind of reproductive advantage, which is not obvious if the entities at issue are human beings. Giere’s verdict (1990) is that evolutionary epistemology is for the moment like Darwinism before the synthesis with genetics, hence it lacks a theory of the mechanisms providing heredity and variation

49 Evolutionary epistemology was not born with Campbell; in fact philosophers like Toulmin elaborated a so-called “evolutionary epistemology”. I consider only recent theories with their massive use of selectionist models.

50 Kenneth Waters (1990) provided a powerful critique of this Darwinian analogy concerning the growth of science, to the extent that it leaves apart intellectual powers of scientists as reflexive agents of selection of fittest contents.
across individuals. However, such an account might be now provided by cognitive sciences.

Like a selectionist theory of culture that is neutral regarding the biological foundations of cultural traits, evolutionary epistemology is not directly committed to any psychological theory of the acquisition of knowledge. There are some evolutionary theories, using selectionist models, that address this point, but this is “evolutionary epistemology” in another sense: “Evolutionary epistemology of mechanisms” (ETM) distinguished by Bradie (1984) from “Evolutionary epistemology of theories” (ETT) of science as a process. They could be complementary (as in Campbell’s Selection Theory (1990)), but they have no logical connection. In general, ETM can nevertheless be understood at an ontogenetical level, or at a phylogenetic level.

Now since this last ETM research program belongs to an evolutionary representation of mind within nature, in order to figure out the philosophical issues at stake, here I now turn to the other strategy, which builds a continuous evolutionary framework for solving questions concerning the nature of man.

2.3.2 Evolutionary psychology

When one talks about evolutionism in psychology, one nowadays could mean many programs, but here I will emphasize a very recent and influential one, namely evolutionary psychology.

The general framework of its strategy is the quest for adaptive value of the features of the human mind. Since the Tooby and Cosmides formulation, the guidelines of the program have been: the human mind is made up of separate cognitive modules that quite unconsciously effectuate successful determinate algorithms that have been constituted through natural selection as adaptations during one of the longest periods of hominization, namely the Pleistocene. One of the main hypotheses is the “cheater detection” module, supposed to solve the problem of discovering the free-riders in situations of reciprocal altruism—situations that should have been frequent in the Environment of Evolutionary Adaptation, and which have been analyzed in evolutionary game theory. Mating behavior and sexual dimorphism51 are an important subject of this research program, within which a robust explanation was provided of some social cognition modules (Cosmides and Tooby 1992), of our computationally amazing

---

51 Papers by Buss and Symons in Barkow et al. (1992), Sterelny and Fitness (2003). A powerful methodological and epistemological critique is given in Kitcher (2003b).
capacities to recognize faces, or of the origin of our ability to represent what others think (called “theory of mind”).\textsuperscript{52}

A large part of that research is conducted in linguistics because new theories of the origin of language become available when people question the adaptive significance of communicating through systems of signs (Pinker and Bloom 1992). Here, the evolutionary framework offers new possibilities in linguistics, among them Chomsky’s idea of a generative grammar embedded in innate competencies acquired through phylogensis, which was incapable of being worked out before one had such a framework.

The major assumptions in evolutionary psychology are that the mind is composed of “domain-specific” algorithms, in opposition to the most pervasive cognitivist hypothesis of some generalist algorithms that are embedded in our minds like some kinds of General Problem Solvers. Here, the psychological thesis fits nicely in the evolutionary framework, because if cognitive competences have been modeled as answers to particular environmental problems, they are necessarily domain-specific. Here, evolutionary theory allows authors to account for flaws or irrationalities in our mind highlighted by the work of Tversky and Kahnemann as a lack of adaptation of our cognitive abilities to our present times (since they were adapted in a very different environment) and also for our quite perfect ability to execute tasks generally unnoticed by traditional cognitivist psychology (recognition of faces, for instance). The last consequence is that cultural traits and institutions must be understood from a prior knowledge of psychological cognitive abilities (Cosmides and Tooby 1989).

Apart from the question of the genic basis of behavior and the commitment of some authors to rather dissatisfying views of genetics (Duprê 2003), evolutionary psychology faces several difficulties: the lack of informations concerning the original Pleistocene environment (which leads to the rise of just so stories),\textsuperscript{53} and the difficulty of testing the main empirical achievements of the research program since competing hypotheses are in circulation.\textsuperscript{54} More generally, the whole program is in need of clarification of what is to be counted as trait, since this decision effects all subsequent empirical investigations. Suppose that we are investigating the evolutionary significance of aggressive behavior: is it the “right”

\textsuperscript{52} Chisholm (2003).

\textsuperscript{53} Lewens (2004) gives a critique of the inference from a reconstituted Pleistocene problem to hypothetical cognitive module.

\textsuperscript{54} Two of the greatest achievements of the program—the Waist-To-Hip ratio theory in the field of mating strategies and the Wason selection task in the field of social cognition—are still severely challenged [Sperber and Girotto (2003) for the second one, with examples of equally corroborated rival theories; Gray et al. (2004) for both of them].
trait to explain? Or is it a part of a general disposition that responds either in a friendly way or aggressively to various situations? Or could it be a too general lumping of various traits: envy, jealousy, territorial ambition, etc.? (Sterelny 2000). Unless this conceptual problem is directly addressed, the whole evolutionary psychology agenda is likely to give birth to diverging, incompatible results with no way to discriminate between them and finally many underlying ideological motives to adopt one or another.

2.4 Conclusions

However interpreted, evolutionary theory is filled with theoretical problems concerning its major concepts (selection, fitness, adaptation) and hence its two major claims of the Tree of Life and of the Hypothesis of Selection. Those problems, albeit never empty of empirical and biological content, are at the same time philosophical since they involve conceptual matters that imply epistemological and metaphysical options. Although they cannot be solved independently of the biological results, and above all could not have been formulated without the acknowledged state of evolutionary biology, they are not likely to be solved only within biological science itself. But, reciprocally, their correct enunciation and the attempt at solutions are of vital interest to the fields of philosophy of science in general and of metaphysics.

However, the most tangible impinging of evolutionary theory on philosophy is the possibility that it gives to elaborate a new framework for many general problems and first of all about the nature of man. I have tried to show the interest of the research programs constituted in this direction, their variety, and the difficulties they are facing. No integrative and synthetic knowledge of man or methodological framework of philosophical problems has yet been accomplished in such an evolutionary spirit that would be parallel and compatible with (and in the end integrated in) the Modern Synthesis. For the moment, we have local results, new challenges never free of ideological and political commitments, and insightful ways of approaching long-standing puzzles (such as the origin of language, maternal attachment, or technological evolution). But in the end, it must be said that this will have profound consequences on the way we generally conceive of philosophical problems and most of all on the image of man that is concerned with those problems.55

55 I thank Elodie Giroux, Françoise Longy, Stéphane Schmitt and Stéphane Tirard for careful reading and suggestions, as well as the editors for their patient revision.
Goldschmidt R (1940) The material basis of evolution. Yale University Press, New Haven
Keller EE, Lloyd E (1992) Keywords in evolutionary biology. Harvard University Press, Cambridge, MA
Lewens T (forthcoming) Seven types of adaptationism in Walsh (in press)
Sober E (1994a) From a biological point of view: essays in evolutionary philosophy. Cambridge University Press, Cambridge, MA
Walsh D (forthcoming) 25 years of spandrels
The Ontogeny–Phylogeny Nexus in a Nutshell: Implications for Primatology and Paleoanthropology

Peter R. Menke

Partout où quelque chose vit, il y a, ouvert quelque part, un registre où le temps s’inscrit.

Henri Bergson

Bestimmt sich nicht auch unsere Erwartung und unsere Bereitschaft, das Neue zu hören, notwendig von dem Alten her, das uns schon eingenommen hat?

Hans-Georg Gadamer

Abstract

This chapter aims to review the relevance of ontogenic data in an evolutionary perspective. Phylogenetic investigation through developmental information is one of the most promising avenues to the elucidation of our natural history. First, the problematic integration of biological subdisciplines into the evo-devo synthesis is considered: the homeobox as Pandora’s box is discussed and the important role of a comparative morphology program is emphasized. Second, the study of development reveals essential aspects of primate supraordinal relationships and does not support an archontan reality. A special note defines the traditional superorder Archonta as (1) an artifact of the Scala naturae concept, since archontans were supreme public servants of the Greek ancient world. On the other hand, it is (2) a vehicle to explain the existence of flying mammals (Chiroptera) via a gliding intermediate stage (Dermoptera). Third, the impact of neotenic ideas on paleoanthropology is retraced and current contributions describing the evolution of the human cranial base and bipedalism are presented. Man’s domination by neoteny seems to be a burlesque, accurately related as pithecocentrism.
3.1 Introduction

The polymath Johann Wolfgang von Goethe was fascinated by the anatomical diversity of the animal kingdom and its variability of form. In 1817 (11 years before Karl Ernst von Baer published his embryological manifesto), he wrote:

- Man findet daher in dem Gange der Kunst, des Wissens und der Wissenschaft mehrere Versuche, eine Lehre zu gründen und auszubilden, welche wir die Morphologie nennen möchten [...] Er abstrahiert bei diesem Ausdruck von dem Beweglichen, er nimmt an, daß ein Zusammengehöriges festgestellt, abgeschlossen und in seinem Character fixiert sei. Betrachten wir aber alles Gestalten, besonders die organischen, so finden wir, daß nirgend ein Bestehendes, nirgend ein Ruhendes, ein Abgeschlossenes vorkommt, sondern daß alles vielmehr in einer steten Bewegung schwanke. ...¹

Life is obviously not a static arrangement but the particular result of a dynamic, ongoing process. This phenomenon is nowhere more relevant than for ontogenetic ongoing studies.

This review approaches some important aspects of ontogeny and phylogeny by presenting a selection of studies that exemplify this scientific field. Embryology, comparative anatomy, and especially molecular biology demonstrate an amazing unity among organisms, and sometimes bizarre variations among them stem from a series of basic themes, some of which are common to all living beings (Nielsen 1995). I concentrate here on broader matters, such as (1) comparative embryology, especially with regard to K.E. von Baer and its modern derivative: evolutionary developmental biology. The ancient realm of morphology plays a major role in this context. By the 1980s, molecular biology and paleoanthropology had developed their own concerns and controversies. Yet the significance of studies going beyond the analysis of adult phenotypes was already necessary and self-evident (Howell 2002).

Since hominin fossils, such as those from Dikika, Taung, Mojokerto, Nariokotome, Tešık-Taš, Kiik-Koba, or Le Moustier, represent infantile or juvenile stages, studying our phylogeny via ontogeny should be of particular importance, although, as Krovitz et al. (2003) pointed out, juvenile fossils are lacking for most species in the

¹ “It can be observed that in the course of art, knowledge, and science, several efforts are made to create and cultivate a doctrine that we may call morphology.... One abstracts the reality of fluctuation by supposing that a belonging together and a fixation of its character can be fulfilled. Considering all shaping, especially in organic forms, we can state that there is no stability, no resting or completing—but rather a fluctuation of all phenomena.” (pp 55, my translation and emphasis.)
genus *Homo*. Here, therefore, studies are reviewed that are based on (2) embryological data sets. Generally, the methodological approaches taken to the analysis of the ontogeny of modern individuals and the fossil record differ strongly and accordingly have different meanings for the study of evolution (MacPhee 1981).

Focusing mainly on current primatological and paleoanthropological aspects, the far-reaching influence of (3) Bolk’s fetalization hypothesis is retraced. Gould’s (1977) interpretation is reconsidered (4), not only based on modern studies but also on traditional work that nowadays maintains only a shadowy existence, although its explanatory power is astonishing. As one may well imagine, any review of the ontogeny–phylogeny question must at present be incomplete. However, *gutta cavat lapidem, non vi sed saepe cadendo* (constant dripping wears away the stone).

### 3.2 Comparative embryology and evolutionary developmental biology

#### 3.2.1 Karl Ernst von Baer and *Gallus domesticus*: the beginnings of comparative embryology

Nineteenth-century comparative embryology lies at the origin of evolutionary developmental biology or, more briefly, *evo-devo*. Von Baer’s and Haeckel’s works are the most popular examples of that period, and these authors’ “laws” describing the general development of organismic form are of great interest.

Karl Ernst von Baer, a pupil of Ignatius Döllinger, a professor at Würzburg, is known as the founder of embryology as a scientific endeavor. Although representatives of German transcendentalism had provided some insights into the field of embryology, it was only in 1828 that von Baer’s *Über die Entwickelungsgeschichte der Thiere: Beobachtung und Reflexion* appeared in print and made even contemporaries recognize him as the founder of embryology. In the first volume of his masterpiece, von Baer concentrated on the development of the chicken (*Gallus domesticus*), but he also bore general laws of development in mind. He worked with dissecting needles and a simple microscope: the “Scholia” describe the deductions he made. The accuracy and minuteness of his fundamental observations is absolutely astonishing. Russell (1982 p 114) was unable to hide his admiration: “His account of the development of the chicken is a model of what a scientific memoir ought to be . . . .”

Von Baer’s ideas incorporate the truly important distinction between the *type* (Wiesemüller et al. 2002) of organization (= the structural plan) and the *grade* of differentiation (modifications of this plan). The aim of comparative anatomy to
reveal group-specific Baupläne now had a prominent new tool by integrating embryological data.

As Charles Darwin was not a professional morphologist, it was others who introduced evolutionary thinking into the realm of comparative embryology. Ernst Haeckel’s formulation of the biogenetic law (ontogeny as the short and rapid recapitulation of phylogeny) presents a radicalized phylogenetic approach. Recapitulatory ideas were not new since Müller (Russell 1982) had already supported a similar hypothesis in 1864. Yet Haeckel’s verve and passion—very striking in a commemorative speech in 1909 (Das Weltbild von Darwin und Lamarck), on the occasion of the 100th birthday of Charles Darwin, favoring Goethe’s monistic world view as being the ultima ratio—made it famous. The interpretations of heterochrony and its implication for paleoanthropology are discussed later on. However, how does today’s evolutionary developmental biology reflect the ideas of von Baer and Haeckel?

3.2.2 The newcomer: evolutionary developmental biology

3.2.2.1 Von Baer and Haeckel—outmoded?

The fancy term evo-devo biology represents an emerging field (Gould 2002; Carroll 2005) that has been specially featured in semi-popular journals such as PNAS (2000) or Nature (2003). It refers to the quest by evolutionary biologists to understand how organisms change shape and form. Hall (2002 p 8) explained that “… evolutionary developmental biology is more than a name for an emerging subfield of biology. It is a reflection of a level of analysis, synthesis, and understanding not possible through the study of evolution or development alone.”

So how are the laws of von Baer and Haeckel interpreted nowadays? Gould (1977) supported von Baer’s explanation and regarded it as being essentially correct while Haeckel’s law is incorrect. Arthur (2002, p 757) disagreed and argued that when comparisons are made between different levels of complexity, a pattern emerges that is broadly (although only in a very imprecise way) recapitulatory. The déjà vu occurs by recapitulating levels of complexity rather than precise morphological details. Haeckel and von Baer are both right in assuming that “evolution leads both to embryonic divergence and, in some lineages, to a lengthening of the ontogenetic trajectory leading to more complex adult phenotypes with greater numbers of cells, their embryos passing through simpler, quasi-ancestral forms.”

Sander (1983) defined a “phylotypic” stage by describing a point of maximum similarity and a succeeding period of divergence. In this scheme, von Baer’s
principle of divergence only applies after this stage. Richardson et al. (1997) revealed a phylotypic period, rather than a stage, by screening more vertebrate species: this supports a broader comparative view. The hourglass model of development (Duboule 1994) might prima facie have reduced the explanatory power of von Baer. We should, however, be aware that this is a different kind of hourglass, the point of constriction being close to the beginning (Richardson 1999).

3.2.2.2 Homeobox-Pandora’s box for the integration of biological subdisciplines?

Lewis (1978), Nüsslein-Völlhard and Wieschaus (1980), and others (see also Lemons and McGinnis 2006) have revealed the meaning of homeobox genes and their importance for investigating the Bauplän. One application of this discovery which causes headaches for many scientists (Kuratani 2003) is the vertebrate cranium: is it a novelty that lies anterior to the head of a prevertebrate organism, or is it homologous to a special region of the amphioxus nerve cord? Holland et al. (1992) recognized a homology based on a comparison of expression patterns of AmphiHox 3 (an amphioxus Hox gene) and Hox-2.7 (the supposed mammalian homolog) and consequently denied the vertebrate brain as representing a novelty of the animal kingdom.

Hillis and Moritz (1990 p 502) emphasized that the power and the rapid development of these techniques “has produced an euphoria in evolutionary biology, because so many new problems can be addressed, it is a commonly held misconception that all evolutionary problems are solvable with molecular data . . .” Nielsen (1995) worried that without morphological characters the molecular data can produce only “naked” trees. Müller’s (2005) review explained inter alia the importance of a comparative morphology program as a prerequisite for an evo-devo synthesis.

Furthermore, Conway Morris (2000 p 4429) warned about “the risk of losing the overall evolutionary context. […] Not only that, but there are intriguing mismatches between genomic architecture and bodyplan complexity.” He pointed out that in all likelihood, a substantial proportion of the metazoan genome was available well before the Cambrian explosion and that a corresponding genetic architecture of genes and bodyplans is required, although he admits that we are still far from understanding either their interconnections or evolution. Arthur (2002) also met the challenge by demonstrating the change of tendency from general laws to specific pathways. He thus called attention to emerging concepts that integrate the overwhelmingly complex data.
As there is no one-to-one correspondence between genotype and phenotype, another important aspect of the problem lies in the interaction of genetic and nongenetic factors that act in different ways upon gene expression and create phenotypic diversity during development and evolution (Hall 1990, 1998b, 2002). Thus begins an appreciation of epigenetics. Goodwin et al. (1983), Raff (1996), and Hall (1998a–c) have called attention to the fact that change in development and evolution must be considered as an interaction of several “biological levels” so that consequently a hierarchical structural analysis is needed.

The timeless stimulus of Carl Gegenbaur—once called the most important morphologist in the world—and his legacy for twenty-first century evolutionary developmental biology lies in his methodological reflections about the relationship between embryology and comparative anatomy. Gegenbaur’s emphasis of careful and detailed observation over tempting and far-reaching speculation should be the *aurea mediocritas* (the golden mean) in our fast-moving world. The modern scientific endeavor faces therefore similar challenges to those of the nineteenth-century evolutionary morphologists, “namely to distinguish between several possible phylogenetic scenarios” (Laubichler 2003 p 29).

As to the formal integration of evo-devo with neo-Darwinian theory not much progress has yet been made. The way toward a unified theory, however, seems to be open, and the stimulation of the biosciences through evo-devo biology is quite evident (von Dassow and Munro 1999; Müller 2005).

### 3.2.3 Primates and supreme public servants of the Greek ancient world

#### 3.2.3.1 Primates and their supraordinal relationships

Although primate origins and supraordinal relationships are discussed twice in Volume 2 (Chapter 1 on morphological and Chapter 2 on molecular data), I would like to mention some studies that have used ontogenetic data in this context. According to Hofer, one of the ultimate goals of primatology (Spatz 1964) is the elucidation of the phylogenetic position of primates compared to other mammals. Gregory (1910) formulated the concept of the Archonta, which unites Chiroptera (bats), Dermoptera (colugos), Menotyphla (Macroscelidea = elephant shrews and Scandentia = tree shrews), and Primates into one single superorder. Later on, modified versions of this hypothesis excluded the Macroscelidea (Novacek and Wyss 1986), while Adkins and Honeycutt (1993)
surprisingly favored a revival of this grouping based on molecular data (COII gene). Even after the Hennigian revolution (1950), many morphologists confirmed the traditional version of primate supraordinal relationships by defining synapomorphies (Hooker 2001). Noncladists, such as Szalay (Szalay and Drawhorn 1980; Szalay and Lucas 1993), supported the evolution and diversification of archontans in an arboreal milieu. The approach is based on the aim of “Darwinian evolutionary classification” to include both adaptive similarity and monophyly sensu Bock and von Wahlert (1965). Critical comments can be found in Grande and Rieppel (1994).

Murphy et al. (2001), however, proposed the new superorder “Euarchontoglires,” based on nuclear and mitochondrial gene sequences of 42 placental specimens. This new grouping (Asher et al. 2005; Nishihara et al. 2006) consists of the Euarchonta (=Dermoptera + Primates + Scandentia) and the Glires (=Lagomorpha + Rodentia). Surprisingly, bats are excluded and do not seem to be closely related to primates. MacPhee (1993 p 372) already noted (based on Adkins and Honeycutt 1993) that “… something is pulling the rodents toward the primates in this data set. . . .”

These radical changes in primate supraordinal relationships consequently deny the Volitantia hypothesis (Illiger 1811), which favors a sister group relationship of bats and colugos (Leche 1886; Thewissen and Babcock 1991; but see Beard 1993). Wible and Martin (1993) documented that the ontogeny of the tympanic floor and roof does not provide any characters distinguishing all extant archontans from other eutherians. Actually, Archonta (archontans were supreme public servants of the Greek ancient world) has Scala naturae written all over it and represents a vehicle to explain the existence of flying mammals (Chiroptera) via a gliding intermediate stage (Dermoptera) in the absence of appropriate fossils (see also Sears et al. 2006). It should be emphasized that the concept of Scala naturae was long developed before the theory of natural selection (Martin 1973). Darwin (1859) promoted the view that nature does not make leaps—Natura non facit saltum—in order (1) to fill the gaps in the fossil record and (2) to strengthen his intellectual position (Schwartz 2000). Although Darwin’s motives are plausible, accepting the Volitantia concept might be an immediate consequence of the emerging battle against supporters of divine creation. Rasmussen (2002 p 7) specified in a more diplomatic manner: “However, it is fair to say that at this juncture we do really not know if primates are more closely related to Scandentia, Plesiadapiformes, Chiroptera or Dermoptera. These four orders are conveniently lumped together as ‘archontan’ in what may be a true clade but which for lack of unambiguous evidence is often used as an informal grouping.” Faute de mieux: chimeric archontans? In the context of primate evolution, some important examples of cranial and postcranial anatomy are presented here.
3.2.3.2 The myth of primate entotympanics: consulting embryological data and its phylogenetic implications

It is now more than a century since embryological studies [the Reichert-Gaupp theory, refuted by Otto (1984)] demonstrated the homology of the mammalian malleus and incus with the articular and quadrate bones which formed the ancestral jaw joint of gnathostomes. This tremendous transition can be traced in fossils by comparing basal synapsids through therapsids to early mammals. A fetal mammal shows that the angular (= tympanic), articular (= malleus), and quadrate (= incus) develop in the same positions they occupy in the cynodont skull (Allin 1975; Maier 1987). The importance of new fossils (Rich et al. 2005), however, is demonstrated by Martin and Luo (2005): the separation of the middle ear bones must have evolved independently among the therians and the monotreme mammals. The middle ear and surrounding basicranium have played a decisive part in mammalian systematics over the past years (e.g. Sánchez-Villagra et al. 2002).

In 1979—2 years before Auditory Regions of Primates and Eutherian Insectivores was published—R.D.E. MacPhee presented a new interpretation of the disputable issue of “primate entotympanics.” Not all mammals show these mysterious skeletal elements. As otic characters are essential for primate taxonomy, this was a potential criterion for distinguishing primates from their relatives [e.g., colugos (Figure 3.1) and tree shrews]. Mivart (1881) defined this specific element as “entotympanic,” while Wincza (1896) introduced the term “metatympanic” (van Kampen 1905). A few influential morphologists (van Kampen

![Figure 3.1](image-url)

Top: Fetal Cynocephalus volans (Dermoptera). Cranial length 41 mm, frontal section. Note the entotympanics. c.E = caudal entotympanic, C.t. = cavum tympani, F.r. = Fenestra rotunda, M.m. = Manubrium mallei, P = petrosal, R.m.a. = Recessus meatus acustici externi, Tu.K. = Tuba auditiva-cartilage, Ty = tympanic, x,y = cartilage. Sections have a thickness of 25 μm. Not to scale. (Modified from van der Klaauw 1922). Center: Late-fetal Cynocephalus variegatus (DUCCE 806). Crown-rump length 136 mm. It shows a frontal section through the promontorium in front of the tegmen tympani. Arrowheads point to branches of the internal carotid nerve running in a transpromontorial position. Scale bar = 0.33 mm. “Cat” = fused cartilage of the auditory tube and rostral and caudal entotympanics, bo = basioccipital, co = cochlea, ct = cavum tympani, e = ectotympanic, eam = external acoustic meatus, pe = petrosal, pr = promontorium of petrosal, t = tympanic membrane. (Modified from Wible and Martin 1993, © Plenum Press). Bottom: The next step. Frontal section of a juvenile Cynocephalus volans-specimen, of 145 mm crown-rump length, from the collection of the Institute of Systematic Zoology in Tübingen. It illustrates the ossification of the fused cartilage of the auditory tube and the rostral and caudal entotympanic (o’“cat”). Hunt and Korth (1980) reported but they did not illustrate the fate of this important anatomical detail. Sections have a thickness of 40 μm. Not to scale
Figure 3.1 (continued)
1905; van der Klaauw 1931) already championed the existence of primate entotympanics.

Van der Klaauw (1922) published an ingenious work (Über die Entwicklung des Entotympanicums) in which he made a strong case for two different types of entotympanic, rostral and caudal. Developing in the anteromedial corner of the membranous tympanic floor, the rostral entotympanic grows posteriorly. Unlike the tympanic processes, entotympanics grow and develop within the fibrous membrane of the tympanic cavity (MacPhee 1979).

Starck (1975 p 143) reported for Tarsius that the existence of an entotympanic is not yet proven. In the younger stage, the tympanic region does not show any peculiarities and an entotympanic cannot be detected. However, an older specimen showed a floor of the middle-ear cavity that is formed by endochondral bone. It was closely connected with the otic capsule, but it showed a rostral extension near the auditory tube: “The ossified bone still contains some cartilaginous tissue, and this would indicate the presence of a true entotympanic, since the petrosal components of the floor are usually formed by periosteal bone. These results are in accordance with the findings of van Kampen (1905).” (My emphasis.) According to MacPhee (1979), primate tympanic floor elements do not develop like any known entotympanic but, rather, like other tympanic processes. Prenatal Microcebus demonstrates secondary cartilage in the rostral tympanic process of the petrosal, which is interpreted as an embryonic adaptation—the explanation of the primordial fusion of once separate entities is no longer required. The hypothesis of a suppressed entotympanic is made more unlikely by the endochondral development of the caudal tympanic process of the petrosal (contra Starck 1975). To quote MacPhee (1979 p 43): “Therefore, with regard to the constitution of the primate ventral wall, and with some necessary violence to Newton’s quip, non fingo ossa - I posit no bones.” Eureka, a new primate characteristic was identified via the careful interpretation of ontogenetic data.

3.2.3.3 Postcranial ontogenetic data and the origin of primates

Stafford and Thorington (1998) and Hamrick (2001) presented important data on the hand proportions in developing and adult mammals. Using a ternary plot (Hamrick 2001), relative metacarpal, proximal, and intermediate phalanx lengths among fossil and extant taxa deliver an elegant possibility of distinguishing primates from other “archontan” mammals. Primates differ from flying and gliding mammals in having much longer proximal phalanges relative to their metacarpals and are unique among the sampled mammals in having elongated proximal phalanges relative to their metacarpals. A comparative analysis of hand
development in the mouse lemur *Microcebus murinus* and other meta- and
eutherian mammals reveals that “interspecific variation in relative digit and
metapodial proportions has high-developmental penetrance; that is, adult
differences are observed at early ontogenetic stages” (Richardson 1999 p 348).
Hamrick’s (2001) results suggest an evolutionary scenario that describes an
“invasion” of the fine branch niche based on a hand with a short palm and
long fingers yielded by a change in digital ray patterning and segmentation.

Recent advances in developmental genetics elucidate patterns of postcranial
growth in primates. Another study is based on a comparison of developmental data
of CD1 mice fetuses and Rhesus macaques: Hallgrimsson et al. (2002) demon-
strated the evidence for morphological integration of the limbs as serially homolo-
gous structures by reporting the covariation structure of forelimb and hindlimb
skeletal elements. This proves that link structures between the limbs are caused by
developmental modules, producing the covariation that would be needed to be
overcome by selection for divergence in fore- and hindlimb morphology.

Since primates have more distally concentrated limb mass than most other
mammals, Raichlen (2005) studied the unique kinematics of primates by testing a
longitudinal ontogenetic sample of baboons (*Papio cynocephalus*). He concluded
that the evolution of primate quadrupedal kinematics was tied to the evolution of
grasping hands and feet.

### 3.2.4 The impact on paleoanthropology

#### 3.2.4.1 Bolk’s fetalization hypothesis: its successors and critics

**Forever young?** Many contributions have emphasized the relevance of phylo-
genetic conclusions from ontogenetic information (Minugh-Purvis and
McNamara 2002). One of the most influential books on developmental change
Tuttle’s (1978 p 287) review of this book was not rhapsodic: “Perhaps the author
attempted too much in this chubby volume [. . .] Time will tell whether it is in
fact a hemicentennial classic as implied on the dust jacket.” Well, in the meantime
it has become probably one of the most frequently cited compendiums. Howell
(2002 p xi) commented that “its effect was immediate, substantive and far-
reaching.” Looking briefly at the bibliographies of modern studies often gives
the impression that the consideration of developmental change in human
evolution began in 1977 (Coqueuniot et al. 2004). Here I focus partly on some
of the “ancient” morphological studies that negate “essentially neotenous” ideas,
to show how profound their explanatory power really is.
Why assume an *impact* on paleoanthropology? The ideas of Louis Bolk, a Dutch anatomist, have in fact had an enormous influence on scientists working in many different fields. The paleontologists Beurlen and Schindewolf (1936, 1950) supported a phenomenon called “proterogenesis” by pointing out that some evolutionary lineages that are present in successive strata may be interpreted as a process of maturing of former embryonic or postembryonic form levels. I mention these thoughts—although they do not deal with anthropological questions—because they show that Bolk’s thoughts are part of a greater, speculative construct of evolutionary ideas. Portmann (1960) already assessed the situation by claiming that criticism of Bolk’s considerations has to be criticism of the entire construct and argued that it is still in progress and may not be considered completed.

As Starck (1962) argued, Hilzheimer (1926, 1927) and others have pointed to morphological and physiological data to explain the principle of fetalization. Starck (1962) traced the hypothesis of neoteny back to Strickland and Melville (1848), but Bolk was the one who applied it to human evolution. A sequence of papers (the version of 1926 being the most frequently cited) takes us away to a non-Darwinian construct. So which are the essential elements?

Bolk (1926 p 5) split human characters into (1) primary and (2) consecutive features. Primary characters are those products of developmental factors that caused the origin of human morphology. Consecutive characters, on the other hand, are phenomena of bipedal adaptation. Hence, the *primum mobile* of human evolution is not bipedalism, the “secondary” characters of which strictly follow functional aspects. Bolk (1926) considered the primary human characters to be: (1) reduction of body hair, (2) form of the external ear, (3) the epicanthic eyefold (=Mongolian eyefold), (4) loss of pigmentation in skin, (5) orthognathy, (6) foramen magnum in a central position, (7) a long persistence of cranial sutures, (8) subcerebral position of orbits, (9) high relative brain weight, (10) position of the spinal column relative to the cranial cavity, (11) women’s labia majora, (12) structure of hand and foot, (13) form of the pelvis, (14) women’s sexual canal in a ventral position, (15) multipapillary kidneys, and (16) the branching pattern of the arch of the aorta [the last two characters are not mentioned by Gould (1977)]. What is the common denominator of these characters?

The characters listed by Bolk are phenomena which temporarily appear during primate ontogeny. Although Bolk (1926 p 44) differentiated the problem by separating (1) the relatives of *Homo sapiens* and (2) the development of human shape, as Slijper (1936 p 504) explained, he advocated the idea that our ancestor must have been an *extant* primate species. Characters of human fetalization consequently represent persisting ontogenetic primate characters. Human
ontogeny therefore demonstrates conservative traits, while humans’ primate ancestors showed “propulsive” (=advanced) characters. Bolk’s (1926 p 26) bottom line is: *H. sapiens* represents a sexually mature primate fetus. However, how did he explain the inhibitive force that fixes man’s physical appearance at a certain point in time?

“The obvious answer is: The slow progress of his life’s course” (Bolk 1926 p 470) and the fact that “human life progresses like a retarded film” (Gould 1977 p 360). Bolk (1926 p 38) asserted that the development of dentition, consciousness, and the late closure of cranial sutures act as indicators of a dominant retardation phenomenon. The chain of causes starts with the modification of the endocrine gland function (for a modern hypothesis see Crockford 2002) by *internal* alteration, not by external factors of the environment (Bolk 1926 p 22).

Slijper (1936), another Dutch scientist, published an outstanding analysis that considered cetacean relationships, the fetalization hypothesis, and the clarification of fundamental terms. Ironically, Gould (1977 p 365) called Slijper’s criticism “famous,” which is quite euphemistic since even the German-speaking Starck (1962) complained about it as not being easy accessible and often ignored. One major problem of Bolk’s idea is the *subjective* splitting of primary and consecutive features. Slijper (1936 p 509) recapitulated Bolk’s scientific career and stated: “...we get the impression that Bolk did not present primary characters (mostly human) at first and discovered their fetal character afterwards, but rather observed a contrarian procedure: he discovered fetal characters and defined them subsequently as primary. This explains the remarkable correlation of primary features with fetal phenomena and diminishes its objective value.” (My translation.) Slijper also denied any general neotenous influence on human ontogeny.

Two prominent morphologists concerned with primordial cranial studies were Benno Kummer and Dietrich Starck. In 1962, they published the first modern study on fetal *Pan troglodytes* (Figure 3.2). Starck and Kummer followed Hofer (1958, 1960) in distinguishing basal and prebasal kyphoses. Gould (1977 p 378) excellently summarized their findings thus: (1) all fetal mammals have a prebasal kyphosis at the junction of the presphenoid and ethmoid bones (a bending with the concave side toward the body, as opposed to a lordosis), (2) this kyphosis decreases during ontogeny, the sphenoethmoidal angle opens out, and the face comes to lie in front of the cranium, (3) while the prebasal kyphosis develops within the basicranial axis during human ontogeny, a different kyphosis develops between the basisphenoid and presphenoid bones at the level of the dorsum sellae. This second kyphosis produces a secondary decrease in the sphenoethmoidal angle following the earlier increase conditioned by straightening of the prebasal kyphosis, (4) the “fetal” value of the
sphenoethmoidal angle in human adults does not reflect the retention of a fetal condition but arises from development of the new, sellar kyphosis. It is a new feature—not a paedomorphic retention.

Amazingly, Gould (1977 p 379) concluded: “These authors have used this single contention as the basis for a campaign against the hypothesis of fetalization . . .” A single contention? A *campaign*? To be blunt, Gould himself started a campaign. In relation to Bolk’s explanation of skull development, Portmann (1960 p 586) already commented that “… skull development is a complex issue which makes Bolk’s solution suspicious.” (My translation.) Reflecting the development and evolution of the human chin, Vogel (1964) warned against too “localized” a view and called for the consideration of the complexity and interaction of all developing skull components. Schwartz and Tattersall (2000) revisited the presence of a chin in hominins and examined the importance of developmental epiphenomena.

Furthermore, Gould (1977 p 379) claimed that the “… tradition of excellence in descriptive morphology is combined with a general avoidance of quantification, and this may have hindered a full assessment.” He might have
overlooked an essential part of Starck and Kummer’s (1962 p 213) paper because “these findings can be characterized as quantitative...”: (1) different relative sizes of neurocranium and auditory capsule, (2) tegmen tympani, (3) frontal nasal region, (4) canaliculus chordae tympanic posterior, (5) commissura orbitonasalis, and others. Gould (1977 p 384), however, did not fall into the Bolkian trap of an all-or-nothing law. Instead, he argued that “most of the classic ‘exceptions’ to human paedomorphosis are really consequences of retarded development,” which Gould described as being the central phenomenon of our heterochronic evolution. Yet his ideas represent an analytic continuation of the Bolkian hypothesis.

Starck and Kummer (1962) did not exclude retardation (as in the commissura orbitonasalis) as an important factor of human development, but they emphasized that accelerations (e.g., the earlier closing of the intermaxillary suture in Homo) as well as deviations (e.g., the basicranium), should influence specific developmental processes. Starck’s (1962 p 23) summary revealed that the principles of human evolution cannot be understood through Bolk’s hypothesis: Important structures of the skull, very often taken as a result of fetalization, are indeed progressive compared with the pongid skull. It is emphasized that the bending of skull base in man and apes is absolutely different, caused by different morphological structures. Identity of causal factors is not established, if we find external similarities, such as the same size of angles, of relative measurements or indices. This may happen by addition of completely different components.

This statement seems not to be a furor teutonicus but rather a well-balanced point of view. Hence, morphometrics might work, but the comparison of morphological details must occur in a correct manner. How do contemporary anthropologists interpret morphological changes of this important cranial region?

**Developmental aspects concerning the evolution of the human cranial base** The cranial base represents the oldest component of the vertebrate skull (De Beer 1985). Thus this “conservative” structure is profoundly important in reflecting man’s phylogenetic history and comparing primates to reveal essential aspects of human evolution. The keystone of the primate skull is definitely the basicranium. Several regions, such as the upper airway, the brain, and other parts, impinge upon one another here and consequently interact during ontogeny (Moss et al. 1982; Dean and Wood 1984; Lieberman et al. 2000). Biomechanically, the cranial base supplies a platform on which the brain develops and around which the face grows. The cranial base also forms a bridge connecting the cranium with the rest of the corpus: providing conduits for all circulatory and vital neural connections, articulating with the mandible and the vertebral column, forming the roof of the nasopharynx, and connecting the sense organs in the skull. Lieberman et al. (2000
p 120) stated that “the shape of the cranial base is therefore a multifactorial product of numerous phylogenetic, developmental, and functional interactions.”

Scientists are confronted with the problem of complicated circumstances in studying this truly important region. It is not only complexity that matters but also problematic ways of measuring. Furthermore, the fragmentary nature of fossil remains causes major difficulties. Novel analytical techniques, however, have helped to engross our thoughts over the past years. Different hypotheses exist that deal with ontogenetic spatial processes and their phylogenetic implications. I mainly follow Jeffery (2003) in reviewing some of the most popular versions. The general spatial-packing hypothesis states that the modern human basicranium is caused by a short cranial base and an enlarged brain. Ross and Ravosa (1993) and Ross and Henneberg (1995) revealed significant positive correlations between increases of relative brain size and cranial base flexion across adult primate taxa: correlation of increasing relative brain size with (1) a coronal reorientation of the petrous bones across extant primates (Spoor 1997), (2) a cranial base flexion using different measurements and landmarks (Spoor 1997; McCarthy 2001), and (3) a cranial base flexion after controlling for the influence of phylogenetic correlations (Lieberman et al. 2000). Enlow and colleagues (Enlow and Hunter 1968; Enlow 1976, 1990) also attempted to demonstrate a determination of cranial base flexion through increases in relative brain size during primate development. Jeffery and Spoor (2002) could not verify these authors’ arguments. They analyzed specimens from 10 to 29 weeks of gestation and documented that petrous orientation and cranial base angulation do not correlate with increases in infratentorial volume relative to posterior cranial base length.

The infratentorial spatial-packing hypothesis has been revitalized by Dean (1988), who argued that having coronally oriented petrous bones and a highly flexed basicranium poses the spatial problem of fitting an enlarged cerebellum on a short posterior cranial base. Jeffery and Spoor (2002) showed that ontogenetic data, collected during the second and early third trimesters of human prenatal development, do not support Dean’s (1988) claim. They indicate that the petrous orientation and cranial base angulation do not correlate with increases in infratentorial volume relative to posterior cranial base length.

The influence on skull form of patterns of brain growth is addressed by two interesting models. Hofer (1969) and Lieberman et al. (2000) favored the brain shape hypothesis, while Ross and Henneberg (1995), Chklovskii et al. (2002), and Sporns et al. (2002) supported a neural-wiring hypothesis. These ideas have in common a suggested necessary change in brain topography to maximize cognitive efficiency by reducing neural wiring lengths. The resulting spatial

Moss et al. (1956) suggested that brain topography is shaped by differential encephalization patterns which lead to developmental changes in posterior cranial fossa morphology.

A few studies (Guihard-Costa and Larroche 1990, 1992; Jeffery 2002) on the human fetal brain showed greater increases in expansion of the supratentorial portion (containing the cerebrum) compared to the infratentorial portion (consisting of cerebellum and brainstem). However, the independence of human cranial base angulation and petrous orientation of changes from the volumetric proportions of the brain between the ages of 10 and 29 weeks gestation are corroborated by Jeffery and Spoor (2002).

Jeffery (2003) tested the key hypotheses by imaging fetal samples of *Alouatta caraya* and *Macaca nemestrina* using high-resolution MRI. He noted marked increases in brain size, especially “disproportionate increases in the size of the cerebrum” (p 281), disproportionate growth of the anterior midline basicranium compared with the posterior midline basicranium, coronal reorientation of the petrous bones, and cranial base retroflexion. Contrary to the spatial-packing hypotheses, increase in relative brain size is not accompanied by flexion of the midline basicranium. Retroflexion is documented for the cranial base in both taxa. There is also little evidence supporting the spatial-packing hypothesis for the fetal period of the howler monkey and macaque due to significant and “seemingly” consistent associations with petrous orientation arise based on background covariations with somatic growth. Jeffery (2003) therefore suggested that laryngeal size might be the reason for basicranial retroflexion. He finally compared it to human fetuses and concluded that the establishment of notable interspecific differences in the basicranium occurs much earlier than in the phase he studied.

Craniofacial growth patterns have been studied by several scientists (Giles 1956; Shea 1983, 1985a, b; Jungers and Hartmann 1988; Ravosa 1991, 1992; Zumpano and Richtsmeier 2003; Cobb and O’Higgins 2004; Mitteroecker et al. 2004). Lieberman et al. (2000) provided a comprehensive review of primate cranial base studies. As Zumpano and Richtsmeier (2003) pointed out, many previous studies documented postnatal growth processes, usually beginning with
growth during the juvenile period. The infant growth period has been incorporated by Ravosa (1992), Richtsmeier et al. (1993), or Shea (1983), while Zumpano and Sirianni (1994) compared fetal to postnatal craniofacial growth patterns. Collections of fetal primates very often do not contain representative specimens (Zumpano and Richtsmeier 2003 p 340). Yet it is desirable to attempt an integration of these stages since only a completely documented ontogeny delivers deeper insight to reveal whether heterochronic processes are responsible for the modifications that have occurred between human and nonhuman primates.

Zumpano and Richtsmeier (2003) investigated, for the first time, growth-related shape changes in the fetal craniofacial region of humans and pigtailed macaques (M. nemestrina), using three-dimensional comparative analysis via cross-sectional samples of CT image data. As they emphasized, a long tradition of studies concentrated on examining the sites of growth of the cranial base, the sites of cranial base flexure, and the determination of the cranial base angle (Bjork 1955; Ford 1956; Dubrul and Laskin 1961; Houpt 1970; Lavelle 1974; Bosma 1976; Lestrel and Moore 1978; Moore 1978; Sirianni and Van Ness 1978; Sirianni and Newell-Morris 1980; Ross and Ravosa 1993; Ross and Henneberg 1995; Zumpano and Richtsmeier 2003). Zumpano and Richtsmeier (2003) showed that decreases in human cranial base length are achieved through the differential growth of posterior and anterior elements. The length of the posterior cranial base decreases, while increases occur in the length of the anterior cranial base. They further argue that a cranial base angle decrease may lead to a total reduction in cranial base length in human fetuses. At a comparable stage, the fetal macaque cranial base does not show a corresponding reduction (increased basicranial flexion). The associated distinctiveness of the differences in midfacial growth and the progression of prenatal cranial base flexion are said to be a factor separating these two species. Zumpano and Richtsmeier (2003) also contradicted Bjork (1955) and Ford (1956) in noting a basicranial flexion—not a constant angle—during the fetal period. They further support Lestrel and Moore (1978), and Sirianni and Newell-Morris (1980) are also supported in assuming a constant macaque cranial base angle during fetal growth, although they report a lesser angle (153°). The human anterior cranial base undergoes more relative growth than the macaque anterior cranial base. For the posterior cranial base, no significant growth differences between these two species are observed. Zumpano and Richtsmeier (2003) speculated that the increases in relative length of the anterior cranial base in humans may reflect the faster rate of growth of the frontal lobes of the cerebral cortex in humans relative to macaques (Enlow and Hunter 1968; Moss and Salentijn 1969; Moss 1973; Sirianni and Newell-Morris 1980) and conclude, based on their own observations and the studies of Anemone and Watts
(1992) and Swindler (1985), that midface differences between humans and macaques reflect a delayed rate of maturation of the human deciduous dentition or an accelerated rate of development. In a tabula rasa manner, Zumpano and Richtsmeier (2003) supported earlier investigations in suggesting the occurrence of shape changes within the fetal craniofacial complex during the last trimester of fetal growth (Grausz 1991; Plavcan and German 1995) rather than assuming an isometric growth process that is, e.g., characterized by size increase without corresponding shape change (Mestre 1959; Houpt 1970; Kvinsland 1971a, b; Lavelle 1974; Moore and Phillips 1980; Sirianni and Newell-Morris 1980; quoted from Zumpano and Richtsmeier 2003). Zumpano and Richtsmeier (2003 p 349) finally concluded that “fetal macaques and humans do not share a common pattern of relative growth of the craniofacial complex, both species undergo increases in mediolateral dimensions (widening) of the skull and increases in palatal and anterior cranial base length.”

One of the ultimate goals in paleoanthropology is to reveal the precise relationship of humans to the great apes, our closest living relatives. Morphological data favor the monophyly of the African great apes, while molecular biology unites humans and chimpanzees (Mann and Weiss 1996; Ruvolo 1997; Enard et al. 2002; Kaessmann and Pääbo 2002). Wildman et al. (2003) even placed chimpanzees within Homo based on molecular data. Pääbo (1999) emphasized the importance of investigating a few genes that are responsible for specific effects during ontogeny (or in adulthood) instead of concentrating on chromosomal rearrangements or the accumulation of point mutations. Hence, Mitteroecker et al. (2004 p 680) stated that “as it is difficult to study gene expression on a molecular level for the whole organism, we confine ourselves to the study of the morphological effects of gene expression during ontogeny.” They therefore created a shape space where each specimen (that is, its landmark configuration) is represented by a single point. In this context, an ontogenetic trajectory corresponds to the ontogenetic sequence which belongs to one species within this space. As Klingenberg (1998) or O’Higgins (2000a, b) showed, geometric contrasts among ontogenetic shape trajectories distinguish the development of different species. Geometric morphometrics is a promising and complex method of collecting and interpreting data based on morphological patterns (Bookstein 1991; Marcus 1996; Dryden and Mardia 1998; Slice 2005).

Some hominid craniofacial growth studies, applying geometric morphometrics, found more or less parallel trajectories from dental stage I (which corresponds to the first permanent molar) to adulthood (Ponce de León and Zollikofer 2001; Penin et al. 2002). The development of hominid cranial morphology consequently diverges from that of the other apes in an early postnatal or prenatal stage. However, O’Higgins (2000a; O’Higgins et al. 2001) confirmed
Richtsmeier et al.’s (1993) assumption of related species subsequently diverging after a similar period of early development. In a comprehensive study, Mitteroecker et al. (2004) measured landmarks and semi-landmarks in relevant specimens following a few days after birth to reveal essential insights into hominid ontogeny. Several principal patterns can be deduced from the set of ontogenetic trajectories. The authors tested three specific hypotheses: (1) “pure heterochrony” of human cranial growth relative to *Pan* is a valid interpretation if the ontogenetic trajectories are identical in shape space, (2) the divergence of human ontogeny corresponds to a similar developmental stage at which the great apes diverge among themselves, and (3) an early divergence of trajectories from common ontogeny could elucidate the considerable morphological differences between humans and great apes because early modifications in development explain drastic transformations of the adult form (Richardson 1999). Studying 206 adult and 62 subadult crania of *Homo sapiens*, *Pan paniscus*, *P. troglodytes*, *Gorilla gorilla*, and *Pongo pygmaeus*, Mitteroecker et al. (2004) collected three-dimensional coordinates of 41 homologous ectocranial anatomical landmarks on the face and cranial base. They demonstrated the expected pattern whereby the youngest specimens are much more similar than the adults (von Baer’s omnipresent discovery). Already at birth, human craniofacial morphology differs markedly from apes (p 692) “in accord with previous studies based on more traditional methods” (Starck and Kummer 1962; Dean and Wood 1984). The first hypothesis, in contrast, can be rejected because there is no sharing of a common ontogenetic trajectory. Penin et al. (2002) tried to revitalize the “neotenic theory” *sensu* Gould (1977 p 365). They do not, however, support a “general, temporal retardation of development” but rather stress that “all the bipedal traits studied, whether in the skull (basicranium) or postcranium (pelvis and femur, see above), do not result from neotenic processes but rather from structural traits” (p 61).

Additionally, the second hypothesis is falsified, and the third hypothesis is supported, by an earlier divergence of the human growth trajectory from the common hominid allometry (Mitteroecker et al. 2004 p 692). The African apes also do not seem to be pure allometric variants of one single type. Mitteroecker et al. thus concluded (p 694) that “pure heterochrony does not sufficiently explain human craniofacial morphology nor the differences among the great apes.” McBratney-Owen and Lieberman (2003) also provide insight into the postnatal ontogeny of facial position in *H. sapiens* and *P. troglodytes* by emphasizing that the ontogenetic integration of complex phenotypes, such as the face, occurs on multiple levels of development, and they further speculate about the effectiveness of ontogenetic analyses for testing hypotheses about natural selection.
Ackermann (2005) investigated similarities in cranial covariation patterns by obtaining measurements from 677 crania of adult and nonadult African apes and sub-Saharan humans to locate underlying developmental and functional causes for the patterning. Defining the points of divergence of the covariation patterns can offer insights into the action of selection on development. Ackermann’s work shows that patterns of integration are similar (not identical) among adult African apes and sub-Saharan humans. Ontogeny documents a sharing of patterns, with each species showing contributions to total integration from the oral region as well as from the zygomatic and to a lesser extent the nasal regions. However, she documented important differences between apes and humans, stating: “In particular, the lower overall integration within and lack of covariance structure similarity among adjacent ontogenetic stages in early human ontogeny differs from what we see in the other apes. It is not entirely clear why this might be so, although it indicates that selection was working in this lineage - either on humans or the apes - to distinguish them not only in morphology, but in variation patterning” (p 195).

Quo vadis? Developmental aspects concerning the evolution of bipedalism I have already mentioned Bolk’s (1926 p 6) interesting ideas concerning human bipedalism “. . . since form became human the posture became upright” (my translation). Summarizing his growth studies on primates, Schultz (1924 p 163) asserted that “man in some respects is less specialized and has hence remained phylogenetically as well as ontogenetically more original and ‘primitive’ than various other primates.”

Structural and mechanical aspects of the locomotion of primates play a considerable role in many discussions of human evolution (Preuschoft 1971; Schaffler et al. 1985; Demes and Jungers 1993; Connour et al. 2000; Ruff 2002; see also Senut, Volume III). Schultz (1953) analyzed over 350 limb bone circumferences and related the results to locomotion. The evolution of bipedal walking has, naturally enough, inspired scientists to associate locomotor mode with the relative lengths of the forelimb and hindlimb bones (Schultz 1937; Napier and Napier 1967; Jungers 1982). Changes within the hominin lineage in the relative size of the upper and lower limb bones are indications of our transition to bipedality (McHenry 1978; Johanson et al. 1982; Wolpoff 1983; Hartwig-Scherer and Martin 1991; McHenry and Berger 1998; Asfaw et al. 1999; Richmond et al. 2002; Ward 2002).

A few French scientists (Berge 1998) have tried to integrate heterochronic processes into analyses of morphological changes during hominid evolution. These works have concentrated on such classic anthropological topics as the anatomy of the pelvis. Berge (1998 p 443) emphasized separating the debate on
neoteny by negating the idea “that identical heterochronic processes occur in skulls and postcranial skeletons, although we know that the growth of cranial and long bones differs in time, rhythm and velocity.” She studied the morphology of two adult pelves and a juvenile hip bone of australopiths, 60 juvenile and adult pelves of modern humans, and 150 juvenile and adult pelves of African apes. The results confirmed a marked difference of the pelvic growth pattern in African apes and humans as reflected in multivariate results, ontogenetic allometries, and growth curves. Two conclusions emerged: (1) a comparison of modern humans to juvenile and adult australopithecines reveals that a unique feature of *Homo* seems to be a prolonged growth in length of hindlimb and pelvis after sexual maturity, while pelvic growth of *Australopithecus* was probably closer to that of apes than to that of humans and that some pelvis traits of adult *Australopithecus* resemble those of neonate *Homo*. Furthermore, (2) at the time of human birth, the appearance of the acetabulo-cristal buttress and the cristal tubercle allows the addition of features, such as the attainment of a proportionally narrower pelvis, with more sagittally positioned iliac blades. In early childhood (as bipedalism is practiced), pelvic orientation and proportions change progressively, while other changes in proportions occur later with the adolescent growth spurt. Neonate *Homo* and adult *Australopithecus* show similar patterns concerning the position of the acetabulo-cristal buttress. This could suggest a later displacement during human evolution. Berge (1998) further documented a progressive displacement of the acetabulo-cristal buttress on the ilium occurring during human growth (from neonate to adult) and hominid evolution (from *Australopithecus* to *H. sapiens*). She finally suggested that evolution of pelvic morphology in hominids is based on a threefold process—predisplacement, acceleration, and time hypermorphosis—and rejected pure fetalization (p 457) by stating that “the present study demonstrates clearly that the concept of neoteny is irrelevant for the pelvis. The study rather implies an accelerated evolutionary process than a retarded one.”

Ruff (2003) examined the human development of femoral to humeral proportions using a longitudinal sample of 20 individuals measured radiographically at semiannual or annual intervals from 6 months of age to late adolescence and also included anthropometric data such as body weights or muscle breadths. A series of limb bone length proportion studies included ontogenetic data (Lumer 1939; Schultz 1973; Jungers and Fleagle 1980; Buschang 1982a; Shea 1983; Jungers and Susman 1984). Ruff (2003) focused on other limb bone dimensions. He compared his results with a cross-sectional ontogenetic sample of 30 baboons. The results document that femoral/humeral length proportions, which are already close to those of adults, are present in human infants, while characteristically femoral/humeral diaphyseal strength proportions only develop after the adoption of bipedalism (at about 1 year of age). Between the age of one
and three, a rapid increase in femoral/humeral strength occurs, and this is followed by a slow increase until mid-late adolescence (when adult proportions are reached). The femoral/humeral length ratio proportions slightly increases throughout growth. There is no apparent growth trajectory change at the initiation of walking and a small decline in late adolescence based on a later humeral growth in length. Also in early childhood, a sex difference in femoral/humeral strength proportions (but not length proportions) develops. Ruff (2003) therefore concluded that they must be largely independent of growth trajectories in strength and length proportions. Baboons (used as a baseline) show contrasting patterns of growth: much smaller age changes in proportions and particularly strength proportions. He therefore stated (p 342): “Comparisons with an ontogenetic baboon sample highlight the specific nature of the human developmental pattern.”

Returning to Adolph H. Schultz, we have an excellent example of a convert. In his youth, Schultz was stimulated by the neoteny hypothesis (see earlier). Following several studies (1953, 1973) although, he rejected Bolk’s idea and the theory of man’s neoteny.

3.2.4.2 Heterochrony in human evolution: a compilation

Repetitio est mater studiorum or as Schwartz (1999 p ix) commented: “... it would be foolhardy to cling unreservedly to a particular set of models and hypotheses without at least occasionally questioning their very bases.” In Sudden Origins, Schwartz reconsidered and discussed inter alia the different interpretation of the Taung child. While Raymond Dart was influenced by ideas of neoteny and concluded that his fossil occupied a missing link-position, Sir Arthur Keith drew a totally different conclusion and defined the Taung individual as an ape. Ironically, Keith’s results were also based on a developmental and even a neotenic chain of ideas (Figure 3.3).

Furthermore, Schwartz (1999) critically reviewed the suggestions of the Australian anthropologist Andrew Arthur Abbie, who equated neoteny and paedomorphism, and portrayed H. sapiens due to its long legs and a highly arched nose as gerontomorphic (=more differentiated from a fetal standard), although he generally described our species as paedomorphic (=little differences from a fetal standard). Abbie absorbed the heterochronic perspective but failed to sustain the argument that human morphological variability is not fixed to a limit and that any combination of features can be incorporated in the same individual.

Gould (1977, 1991 p 2) called attention to the difference between Haeckel’s heterochrony, which describes the pathway of development in an organ relative to
the pathways of other organs in the same animal, and current usage, which defines it as “the course of a trait relative to the ontogeny of the same trait in an ancestor (or related form).” Jena’s most famous zoologist was not really interested in elucidating evolutionary mechanisms as a method to explain the patterns of character states. He was rather interested in reconstructing natural history by applying biogenetic laws. Sir Gavin de Beer (1930) opened up new vistas by modifying Haeckel’s concept: any change in the timing of a character is compared to the same feature in an ancestor. This established the study of heterochrony within the modern synthesis (Gould 1991).

McNamara (2002 p 1) defined heterochrony essentially as “change to the timing and rate of development” which produces two major effects: *paedomorphosis* (if there is less growth during ontogeny, the descendant adult will resemble the juvenile condition of the ancestor) and *peramorphosis* (where the descendant undergoes greater development). Three different processes generate paedomorphosis and peramorphosis, respectively: (1) progenesis (prematurely truncated duration of growth in the descendant), neoteny (a lesser growth rate in the descendant than in the ancestor), and postdisplacement (delayed onset of growth) and (2) hypermorphosis (extension of duration of growth in the descendant), acceleration (an increase of growth rate in the descendant), and predisplacement (earlier onset of growth in the descendant). The whole organism can only be affected by progenesis and hypermorphosis, while the remaining aspects affect certain traits of the organism. McNamara (2002) further stated that peramorphosis and paedomorphosis are the products of varied processes rather than being processes in themselves. Hypermorphosis (extensions of the end of growth) and progenesis (contractions of the end of growth) are considered within a global context, “based on changes to the time of onset of sexual maturity and cessation of somatic growth, with the two frequently coinciding” (McNamara 2002 p 105).

Figure 3.3
The Taung child (left), compared to the skulls of comparably aged gorilla (middle) and chimpanzee (right). Dart’s desire was to show that the apes had begun to grow away from the juvenile state. The Taung specimen has no brow ridges, small jaws, and a very large brain (modified from Schwartz 1999, drawing by T. D. Smith, © J. H. Schwartz)
Different views have emerged on how large a role neoteny plays. While Gould (1977), Antón and Leigh (1998), and Montagu (1989) supported true paedomorphosis (neoteny), Shea (1988, 1989, 1992), McKinney and McNamara (1991), and McNamara (1997) favored some kind of hypermorphosis in time. Bogin (1997) suggested a new developmental stage in between (no heterochrony); Alba (2002) on the other hand, characterized paedomorphosis and peramorphosis as interpretative, not descriptive, terms. He emphasized the importance of modifying conventional clock models based on meaningful variables. Alba (2002) called for a double standardization (initial and final developmental stages) of ontogenetic trajectories and suggests a “complete” model including absolute age (at homologous developmental stages), shape, size, and behavior. As he admitted, this is not an easy undertaking.

McKinney and McNamara (1991) have indicated that, during ontogeny, contractions and extensions can occur at transitions between particular life-history stages and that local growth fields can also be modified. The reality that heterochrony may operate at any time during ontogeny (from the point of fertilization until the cessation of growth) has been blurred by the comparison of the cessation of a descendant’s growth with the ancestral condition (McNamara 1983; McKinney and McNamara 1991; Maier 1999). Mammals, for example, are characterized by embryonic and postembryonic (infantile, juvenile, adolescent, and adult) growth phases. In his thought experiments, McNamara (2002) (Figure 3.4) showed how some authors have described paedomorphosis where it was nonexistent. Sequential hypermorphosis (defined by him in 1983 as terminal hypermorphosis), for instance, also effects the offset of growth and therefore implicates greater development within each growth stage based on “scaling effects and probable increase in size, either of the part, or of the whole” (2002 p 108). Hence, the descendant is compared to the ancestor in a relatively more juvenile state. Paedomorphosis, however, should be defined by adult characteristics.

What kind of consequences does sequential heterochrony have for our understanding of human evolution? The list of “general neoteny—supporters” is long (I have already mentioned Bolk and Gould, but see also Montagu 1989; Wolpert 1991). Montagu’s (1989) favorite ancestor is a form very like the pygmy chimpanzee in order to compensate the incorrect assumption of Bolk that the hominid line has passed through an ape-like stage such as the gorilla–orang type. Provided that human is “essentially neotenous”, hominid evolution would have produced an organism characterized by having a smaller body size, brain, and limbs (Shea 1989; McKinney 1998). McNamara’s (2002 p 115) summary described hominid evolution as involving “a mixture of peramorphic and paedomorphic
McNamara’s (2002, © Johns Hopkins University Press) representation of the ontogeny of ancestral and descendant species that pass through distinct growth stages. (A–C at times T or T′). (a) A greater sequential hypermorphosis results in failure of the descendant to pass into stage C before cessation of growth at time T3. (b) Sequential hypermorphosis (Seq Hyp) at A and B results in the period spent in stage C being much shorter in the descendant. (c) Terminal hypermorphosis (Term Hyp) and sequential hypermorphosis result in all descendant growth stages being extended. (d) A shows an extreme sequential hypermorphosis and results in stage B being omitted in the descendant.
traits.” Forty years back, Starck and Kummer (1962) had come to a similar conclusion via a totally different method.

Shea (2002, p 95) applied his working hypothesis—size diversification occurs via predominant rate changes (rate hypomorphosis and hypermorphosis) rather than by time changes (time hypomorphosis and hypermorphosis)—to human evolution, and concluded that: “Certainly, at present no emergent data support any genetic or developmental basis for a global or generalized neoteny.” He accused Godfrey and Sutherland (1996) of revitalizing the idea that hominid evolution has predominantly involved a generalized neotenic transformation. This accusation is problematic in that Godfrey and Sutherland (1996 p 40) freely admitted that “Gould’s neoteny hypothesis for human evolution has been criticized on a number of grounds. The thesis of this paper is that both Gould and his critics overstated their cases: Nothing that we have said should be construed as a defense of Gould’s hypothesis. Our intention is simply to reopen the dialogue, and to propose a framework for more precise testing of heterochronic hypotheses” (My emphasis).

In contrast, Hall (2002 p 13) attacked the one-sided concentration on heterochrony as the only way to consider development and evolution by claiming that it “seemed that everyone could find evidence for heterochrony or at least justify use of the term to explain phenotypic changes in their favourite organism. Other mechanisms linking development and evolution were ignored or not sought.” He emphasized the importance of heterotopy, the spatial pendant of heterochrony.

In summary, I give below some important terms and definitions (see also earlier) based on Minugh-Purvis and McNamara’s (2002) summary, to present a brief synopsis:

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acceleration</td>
<td>A heterochronic process that involves a faster rate of development in the descendant and also produces a peramorphic trait in the descendant</td>
</tr>
<tr>
<td>Deceleration</td>
<td>A slower rate of developmental events in the descendant. It produces a paedomorphic trait when expressed in the adult phenotype (syn. neoteny)</td>
</tr>
<tr>
<td>Hypermorphosis</td>
<td>Developmental events show a delayed cessation (or offset) in the descendant. Expressed in the adult phenotype, peramorphic traits are produced. Global hypermorphosis can be caused by late sexual maturation (terminal hypermorphosis). But hypermorphosis can also be caused by a delayed cessation in local growth fields</td>
</tr>
<tr>
<td>Isomorphosis</td>
<td>Peramorphosis followed by paedomorphosis (or vice versa). The descendant does not show any effective morphological changes. Unusual phenomenon of paedomorphosis</td>
</tr>
<tr>
<td>Neoteny</td>
<td>A slower rate of developmental events in the descendant, producing paedomorphic traits when expressed in the adult phenotype (syn. deceleration)</td>
</tr>
</tbody>
</table>
Paedomorphosis: The descendant adult retains subadult ancestral traits
Peramorphosis: The ancestral adult shows development of traits beyond the “usual” ancestral stage
Progenesis: Developmental events show an early cessation in the descendant. Expressed in the adult phenotype, peramorphosis is produced (syn. time hypomorphosis)
Sequential heterochrony: Contraction or prolongation in the descendant relative to the ancestor of life-history stages or ontogenetic growth
Sequential hypermorphosis: Life-history stages or ontogenetic growth are prolonged in the descendant relative to the ancestor (syn. proportional growth proportion)

3.3 Conclusions

Alberch (1982) rightly pointed to the importance of developmental constraints in evolutionary processes. It is nevertheless also important to emphasize the highly informative aspects of development (e.g. Britz and Bartsch 2003; Cracraft 2005; Alemseged et al. 2006; Brakefield 2006)—we should appreciate the synergistic effects. Considering paleoanthropology Dullemeijer (1975 p 86) predicted that “the fundamental idea can inspire many in their future research if the foetalization theory is reduced to the right proportions, ... It seems to open great perspectives for anthropology.” Two years later, Gould (1977) responded to this challenge and inspired a whole generation of biologists.

Most studies overwhelmingly reject a general neoteny phenomenon in human evolution. The conclusion of man’s domination by neoteny represents a pitheco-centric perspective: a fruitful, but wrong hypothesis. Sophisticated techniques, such as geometric morphometrics, and more embryological data indeed improve the quality of the data sets, but on the whole often verify traditional contributions. Gould’s (1977) brilliant forte, however, was primarily to break down the language barrier and to revitalize the topics of ontogeny and phylogeny. Summa summarum, I would like to conclude with Dullemeijer’s (1975 p 84) take-home message—not only in the figurative sense—by stating: “Homo sapiens has its own face.”

Acknowledgments

I am very grateful to Winfried Henke, Hartmut Rothe, and Ian Tattersall for inviting me to contribute. For intellectual support, I would like to thank Winfried Henke. Special thanks to Ian Tattersall for improving my pseudo-English manuscript, and I am very grateful to Jeffrey H. Schwartz for important comments. For inviting me to study Cynocephalus volans and for enriching discussions thanks to Wolfgang
Maier from the Institute of Systematic Zoology in Tübingen. Thanks also to Gerhard Storch, John R. Wible, Christian Kummer SJ, and Thorolf Hardt. The postgraduate financial support of the Johannes Gutenberg-University Mainz is gratefully acknowledged. Finally, Dankescho¨n to all my loved ones.

References


Bolk L (1926) Das Problem der Menschwerdung. Fischer-Verlag, Jena
Haeckel E (1923) Das Weltbild von Darwin und Lamarck. Alfred Kröner Verlag, Leipzig


Illiger C (1811) Prodromus systematis mammalium et avium additis terminis zoographicius utriusque classis, eorumque versione germanica. Berolini


Lumer H (1939) Relative growth of the limb bones in the anthropoid apes. Hum Biol 13: 283–305
Mivart G (1881) The cat. An introduction to the study of backboned animals, especially mammals. Scribner’s, New York pp 557
Schindewolf OH (1950) Grundfragen der Paläontologie. Schweizerbart, Stuttgart

The ontogeny–phylogeny nexus in a nutshell: implications for primatology and paleoanthropology


Starck D (1962) Der heutige Stand des Fetalisationsproblems. Verlag Paul Parey, Hamburg und Berlin


Strickland HE, Melville AG (1848) The Dodo and its kindred or the history, affinities and osteology of the Dodo, Solitaire and other extinct birds of the islands Mauritius, Rodrigues and Bourbon. Reeve, Benham and Reeve, London 141 p


Van der Klauw CJ (1922) Über die Entwickelung des Entotympanicums. Tidjschr Ned Dierkdv Ver 18: 135–174


Von Baer KE (1828) Über die Entwickelungsgeschichte der Tiere: Beobachtung und Reflexion. Bornträger, Königsberg


4 Principles of Taxonomy and Classification: Current Procedures for Naming and Classifying Organisms

Michael Ohl

Abstract

Taxonomy deals with the naming and classification of organisms and is an integrative part of biological systematics, the science of biodiversity. The information provided by taxonomic research is a fundamental basis for all fields of biology. Current taxonomy focuses on multicharacter integrative approaches, considering all potentially useful sources of information provided by the various fields of biology. The resulting supraspecific classification should be based on the genealogy of organisms, that is, on a phylogenetic analysis, to be objectively testable. However, for pragmatical reasons, a classification based on overall similarity and diagnostically relevant characters might be a heuristically important step in taxonomy and should be perceived as an approximation to a classification tested by phylogenetic methods. The nested levels in a classification of organisms are usually not only named but also ranked, that is, a set of hierarchical terms like genus, family, and class, are applied to reflect the hierarchical structure of the classification. Assigning these so-called Linnaean categories to a classification is (1) a voluntary action to make a classification notionally more easy to access; and (2) a linguistic activity that is done subsequent to obtaining the scientific results of the systematic analysis.

4.1 Introduction

Taxonomy, the science and method of naming organisms, is a fundamental basis for all biological science and its application. The primary task of taxonomy is to describe, establish, and give an account of the order that is an inherent property of biological diversity. The order of names provided by taxonomy is arranged as a hierarchical classification, which is considered to portray the hierarchy of species and more inclusive taxa as a result of the continuous chain of species splittings in
the evolutionary history of life on earth. Generalizations on organisms as a basic principle in biology are only possible if the infinite number of items in science is classified. Statements about the overwhelming diversity of nature would be impossible without methods for bringing order to this diversity. The world’s biota is a vast library of information concerning any aspect of life, and taxonomy is the cataloguing system that everybody must use to access its information. All kinds of biological science and applications link their specific data to species names and use these names for effective communication. As Longino (1993 p 85) has paraphrased, “… taxonomy is the raw material from which hypotheses of phylogeny are derived.” All kinds of comparative biology rely on sound phylogenetic hypotheses, and the reliability of a phylogenetic hypothesis immediately depends on the reliability of the underlying taxonomic data. Moreover, society has an increasing need for reliable taxonomic information in order to allow us to manage and understand the world’s biodiversity (Wheeler et al. 2004). Until recently, taxonomy was confronted with what Godfray (2002 p 17) called a new bioinformatics crisis, evidenced “by a lack of prestige and resources that is crippling the continuing cataloguing of biodiversity.” Current biological taxonomy quite successfully adopts methods, data structure, and other demands of techniques and theories invented by new entrants to the biological sciences such as the fields of molecular biology (e.g., DNA barcoding; see DeSalle et al. 2005). However, all other useful sources of information are simultaneously gathered in modern taxonomy, and this multicharacter integrative approach has been called integrative taxonomy (Dayrat 2005; Wheeler 2005). It allows taxonomists to create new common visions to meet changing demands of a changing global view on global biodiversity and the threats to it.

### 4.2 Definitions

Inconsistent terminology is a barrier to communication, results in confusion and misunderstanding, and prevents effective science. The variation in definitions for “taxonomy,” “systematics,” “classification,” and related or derived terms is as complex as it is contradictory. The major problem with these obviously closely allied terms is to differentiate them from each other, and many authors treat them more or less as synonyms, either intentionally or intuitively. In a recent textbook on biological systematics, Schuh (2000) implicitly equated systematics, classification, and taxonomy, when he defined systematics as “the science of biological classification.” The opposite view is held by Hawksworth and Bisby (1988 p 10), who suggested restricting taxonomy to “taxonomic information systems (classification, nomenclature, descriptions, identification aids)” and defining systematics
to include “taxonomy in the above restricted sense together with analyses of variation, of phylogeny, evolutionary processes, etc.”

Probably most current scientists would agree that the most appropriate name for the scientific area composed of taxonomy, systematics, classification, and all of their respective concepts, theories, and methods is “biological systematics,” or systematics in short.

What are the basic units of systematics? Systematics is not concerned with individual organisms, although these are always studied by systematists as representatives of species or other supraspecific groupings. The simplest and most descriptive statement would be that systematics deals with taxa. What then is a taxon (plural, taxa)? Simpson (1961 p 19) defined it as “a group of real organisms recognized as a formal unit at any level of a hierarchic classification.” This definition recognizes groups of organisms as taxa only if they are already formally classified, which is not always necessary or wanted. Newly discovered groups, hypothetical groups that still await confirmation, putatively artificial groups that are still disputed, and many others do not qualify as taxa under Simpson’s definition. In the field of phylogenetic systematics or cladistics, the taxon definition is often narrowed even further to monophyletic groups of organisms (Sudhaus and Rehfeld 1992; Mayr and Bock 2002). The final objective of systematics is indeed to include monophyletic groups only, but for a wide variety of reasons this goal can hardly be achieved right away. Hypotheses on monophyly are created and rejected, groups prove to be nonmonophyletic but are still being discussed, putatively monophyletic groups are still of unknown position within a certain higher lineage. Other problematic cases are fossil stem groups, like Australopithecinae, which are obviously paraphyletic with respect to Hominidae but can be treated as a heuristically important grouping in paleoanthropology. All these groupings are best referred to by a generalized term, and taxon is the most appropriate one. Hence, a definition of the term taxon might be: “A taxon is a group of organisms that can be differentiated from other groups of organisms, and that can be described and named.” Other terms to replace the term taxon have been proposed that are considered to carry specific connotations to reflect a specific framework for which they have been proposed. Terminal taxon, as used in cladistics, and operational taxonomic unit (OTU), proposed by the phenetic school of systematics, are two prominent examples that have been suggested to designate the units of systematics without reference to a particular rank or phylogenetic position. However, the term taxon, as here defined, is sufficient to refer to ranked and unranked groups of organisms.

Biological systematics in fact deals with taxa, that is, biological populations, species, and higher taxa. However, systematics is concerned not only with the taxa
themselves, their description and characterization but also with their origin, development, and other kinds of interrelationships. The fundamental and underlying concept of systematics is evolution, and Huxley (1940a, b) regarded systematics as “detecting evolution at work.” Simpson’s (1961 p 7) classical definition of biological systematics (or just systematics) is still adequate: “Systematics is the scientific study of kinds and diversity of organisms and of any and all relationships among them.” This definition was subsequently simplified to include the modern term “biodiversity,” e.g., Mayr and Ashlock (1991) and Sudhaus and Rehfeld (1992 p 11): “Systematics is the science of the diversity of organisms.”

Another term that plays an important role in biodiversity research is taxonomy. Taxonomy is not easy to differentiate from systematics, and it has been frequently intermingled with systematics. If separated, taxonomy usually refers to what Hawksworth and Bisby (1988 p 10) called “taxonomic information systems,” i.e., naming of taxa, nomenclature, descriptions, identification aids, whereas systematics is widely understood in a general sense of including taxonomy and the analysis of phylogenetic relationships, speciation processes, variation, and others. Since under any definition the terms taxonomy and systematics are closely associated, many authors have simply synonymized them because they consider their separation not to be feasible. The development of a concise methodology and theory of phylogenetic systematics in the last decades has considerably influenced the practice of modern taxonomy. Taxonomy, as understood here, might be treated as a field of systematics, with specific methods, theories, conventions, and terminologies that are different from other systematic fields (Will and Rubinoff 2004). There is, of course, continuous, active transgression of the borders between taxonomy and other branches of systematics, such as phylogeny and biogeography, and the most desired taxonomic revisions are obviously those whose taxonomic decisions are tested by phylogenetic analysis. To summarize, an appropriate definition of taxonomy might be (modified from Winston, 1999 p 9):

- Taxonomy is the branch of biological systematics that is concerned with naming of organisms (according to a set of rules developed for the process), identification (referring specimens to previously named taxa), and classification (ordering taxa into an encaptic hierarchy based on perceived characters).

The set of rules that governs the formation and use of taxon names in zoology is provided by the International Code of Zoological Nomenclature (hereafter referred to as the Code), published by the International Commission on Zoological Nomenclature and currently in its fourth edition. The Code and its bearing on taxonomy will be briefly discussed later. The definition described earlier points out that a classification of organisms is based on perceived characters. Any organism consists of innumerable numbers of characters that can potentially
be perceived by a biologist, but not all of them qualify as characters helpful in achieving a meaningful classification. The so-called phylogenetic systematics provides the objective framework for evaluating characters according to their meaning for the evolution of a taxon. The most reliable and testable classification is based on the results of a phylogenetic analysis, which tries to reconstruct the hierarchical structure of organisms as a result of the continuous sequence of speciation events in evolution. This should be the basic goal of all taxonomy. However, for different reasons, phylogenetic relatedness might (still) be unknown, and thus the set of perceived characters for establishing a sound classification might be selected inter subjectively by means of overall similarity and diagnostically relevant characters. A classification based on overall similarity might be a heuristically important step in biological systematics and should be perceived as an approximation to a classification based on phylogenetic hypotheses. In practice, such non- (or pre)phylogenetic classifications might stand for a long time, but they should be regarded as preliminary until tested by phylogenetic methods.

The previously mentioned definition of taxonomy implies that one of its main goals is a classification. As paraphrased by Schuh (2000 p 15), a “classification represents the codification of the results of [bio]systematic studies.” A general definition of classification might be

- A classification is a set of names that are ordered hierarchically and more and more inclusive. The hierarchy of names is considered to portray the hierarchy of organisms as the result of the evolutionary process.

Again, the hierarchy of organisms should preferably be recovered by means of phylogenetic methods, but might be preliminarily but meaningfully approached by the criterion of overall similarity.

Some authors (Griffiths 1974; Wägele 2005) differentiate classification from systematization for fundamental reasons. The term classification originates from the term class as in use in the philosophical logic and means a group of objects that have a specific set of properties in common. Which property is selected as class-defining is purely subjective and arbitrary, and a classification based on this property is similarly subjective. Thus a classification, as defined by Wägele and others, is considered to be a strictly conceptual system and is the product of an arbitrary, linguistic activity. In contrast, organisms are the product of the historical process of evolution, which is expressed as a continuous chain of speciation events during the history of life. The sequence of species splittings over millions of years has resulted in a hierarchical structure of the organisms that systematists achieve to reconstruct. Thus, this kind of hierarchy is conceptual as well, but it is considered to represent the hypothetically real sequence of speciation events in evolution. This process is called systematization by Wägele (2005).
The conceptual representation of the mental order as achieved by the process of systematization is called a phylogenetic system. The difference between a classification and a systematization is of fundamental importance but is usually not considered in practical biological systematics. I use the term classification for both ordering systems here, which is in accord with the widely accepted understanding, but we must keep in mind the resulting double meaning of classification and the respective subtle differences to systematization.

A completed classification is a hierarchy of names that denote hypotheses on taxa. It is organized in a more and more inclusive terminological structure of a theoretically infinite number of levels. These levels may bear names, but many of them, particularly in fully resolved cladograms with vast numbers of levels, remain unnamed for practical reasons. After completing a classification and selecting an appropriate number of higher taxa to be named, the hierarchically organized set of named levels can be given relative ranks as individual designations. Such relative ranks are provided by categories, such as classes, families, and genera, and this system of categories is well known as the Linnaean hierarchy or the Linnaean categories. Although the meaning and the current status of the Linnaean categories will be discussed later in more detail, it should be stressed here that an assignment of Linnaean categories to the level of a classification is (1) a voluntary action in order to make a classification notionally more easy to access; and (2) a linguistic activity that is done subsequent to obtaining the scientific results of the systematic analysis.

As we have seen previously, names as linguistic elements are of crucial importance in biological systematics. Names serve as labels to relate to concepts and hypotheses about taxa and their evolution. Names should principally be unequivocal, universal, international, and perpetual. Due to the incommensurably large number of organisms on earth, not to speak of the immense numbers of theoretically possible higher taxa that all could, at least in principle, be named, the need for internationally binding regulations is obvious. The part of taxonomy that is concerned with assigning names to taxa is called nomenclature. The sets of regulations that govern the practical application of nomenclature are provided by the International Code of Zoological Nomenclature.

### 4.3 Taxonomy as a system of ordering data

Biologists deal with a remarkable diversity of items. For any scientific procedure in biology, it is of critical importance to differentiate and thus to compare these items. It is one of the intrinsic attributes of a comparative process to arrange the to-be-compared items according to specified properties, that is, to classify these
items. Mayr (1995) differentiated four different kinds of classifications: (1) special purpose classifications (based on particular features of special importance in the context; e.g., ecologists may divide small mammals into soil-dwellers and tree-dwellers); (2) downward classification by logical division (starting with the entire animal kingdom, groups are defined on the basis of dichotomous splittings in two less inclusive groups, such as the warm-blooded and the cold-blooded animals, until the species level is reached); (3) upward or grouping classification (items are arranged in more and more inclusive groups according to observed characters); and finally (4) Hennigian phylogenetic or cladistic system (grouping of species and higher taxa on the basis of common descent). With respect to their nature as ordering systems, the upward classification (3) and phylogenetic systematics (4) do not differ, because in both classifications species are combined into higher taxa (ranked as genera), these higher taxa are combined into even more inclusive taxa, and so forth. However, phylogenetic systematics is a special kind of upward classification in permitting only one criterion as the basis for the classification, that is, common descent.

Five basic objectives of ordering systems can be differentiated in general and apply to biological systematics as well (Vane-Wright 2001; Mayr and Bock 2002): (1) discrimination (delimiting groups against other groups); (2) information storage and retrieval (the structure of classification systems, based on different criteria that are considered to be informative regarding the group, permits storage and retrieval of a large amount of information); (3) recognizing group affiliations (this is called identification and refers to the process of referring an organism to a previously described taxon, e.g., by the use of a dichotomous key); (4) inferences about not yet studied properties (it is a widely held prediction that many of the characters of a taxon studied will be similar or even identical to those of closely or immediately related taxa); (5) to serve as a baseline in comparative studies (the validity of comparative studies in biology largely depends on the reliability of the preceding research by which the studied items were grouped). These are basic objectives of that part of biological systematics that focuses on the biological items themselves, i.e., organisms, species, and higher taxa. The scientific treatment of these items requires a scientific terminology to communicate, which is provided by biological nomenclature.

4.4 Taxonomy and classification without phylogeny: an outdated remnant or a practical necessity?

There is an increasing shift in taxonomy toward modifying the directives on how to name organisms to reflect genealogical relationships. However, despite the fact
that most systematists would agree that the hierarchical system of names in biology should be governed by phylogenetic hypotheses instead of by intuitive similarity, the increasing pressure of phylogenetic demands on the naming processes results in intrinsic conflicts in systematics. Phylogenetic analyses have not been conducted for most groups of organisms yet, and it is highly unlikely that many will be undertaken soon. Additionally, new species are rapidly discovered in large numbers, particularly among invertebrates. Since a taxonomic treatment of such new taxa, that is, formal descriptions in taxonomic revisions or even single taxon descriptions, is a prerequisite to any other scientific exploration, new taxa are more rapidly described and made available than phylogenetic hypothesis are worked out. The taxonomic treatment of certain taxa, such as those that are less known, newly discovered, and or extraordinarily species rich, is a fundamental first step that opens up the possibility of continuing with subsequent phylogenetical, genetical, behavioral, and other studies.

In practical taxonomy, it is often necessary or at least recommendable initially to leave a comprehensive revision or a phylogenetic reconstruction aside from research projects and the resulting publications. In principle I agree with Mayr and Ashlock (1991 p 347) that “the isolated description of . . . new species . . ., divorced from revisional or monographic work, is the least desirable form of taxonomic publication.” But Mayr and Ashlock correctly qualify their assertion themselves in stating that it does not hold “in well-known groups.” A well-known group is usually a taxon which is not only well known in terms of scientific research but also which additionally receives extraordinary attention by both the scientific and the general public. It is thus not surprising that isolated descriptions of fossil and recent Primates and large mammals often make their way to the highest ranked scientific journals (e.g., the description of *Homo floresiensis* by Brown et al. 2004). Obviously, the publicity that results from such publications can be a strong motivation for any scientist to publish a rather isolated description immediately rather than to invest more time to obtain more data for a more comprehensive publication, due to the current system of scientists being under considerable pressure to publish and to compete for limited research money and, finally, jobs.

There are several more potential reasons to conduct descriptive taxonomy (Godfray 2002), which may mutually affect each other in most cases. Examples are: (1) it might be desirable to make a new discovery formally available for further studies, which can or should not be conducted at the time of discovery of the new taxon for varying reasons; (2) if a scientist has discovered a new species, and if he intends to describe it formally, this scientist might want to guarantee that the name of this taxon as proposed by him is the first formal description and, thus, has priority over any subsequent name. This might be considered as to
imply a connotation of personal rather than true scientific motivation, similar to the argumentation as presented earlier. However, priority is a basic principle of zoological nomenclature, and although it is explicitly thought to be priority of publication without reference to the date of discovery, the publication date here serves as an objectifyable reference point. Hence, priority of publication just replaces the priority of discovery for practical reasons; (3) incomplete knowledge of data might prevent systematists for conducting a comprehensive revision or a cladistic analysis. This might be due to the lack of characters (e.g., in fossils or other incompletely preserved specimens) or material (e.g., modern techniques such as molecular systematics or scanning electron microscopy require fresh or specifically preserved material, but many rare but potentially informative species are known only from dry museum material, which cannot be fully examined).

Another topic concerns the role of biological systematics in the context of a global biodiversity assessment. The development of global species inventories is considered to be an urgent and vitally important task that is a primary step and fundamental activity for any kind of biodiversity research (Stork and Samways 1995; Purvis and Hector 2000; Wilson 2003). Although it seems to be clearly unrealistic to describe every species of organism on earth, not to speak of the monumental uncertainties as to how many species there are (Godfray 2002), any step toward a global inventory of selected “target taxa” should be achieved as soon as possible.

Phylogenies are inherently hypothetical, simply because they portray historical processes, which cannot be inferred directly. Hypotheses, however, can be of different quality. The reliability of a hypothesis largely depends on the quality of the underlying data, which involves numerous theoretical and methodological aspects. Among these aspects, the completeness of the data set is of crucial importance. Completeness of characters is an illusion, since each single organism theoretically consists of an infinite number of characters. Completeness of taxa is, at least in principle, possible. With regard to taxon sampling, the perfect systematic study would include all species of a given taxon, perhaps even both still-living and extant. Practically, completeness of taxa is unlikely to be achieved in most groups, except perhaps for some exceptionally well-studied taxa such as birds. However, the reliability of a phylogenetic analysis increases with the increasing completeness of the taxon sampling. Conversely, gaps in the taxon coverage result in considerably less reliable phylogenetic hypotheses. In many cases, phylogenetic analysis should better wait for a more complete species inventory, which, particularly in taxa with many species or with a patchy distribution, can hardly be achieved in one step.

In summary, publications concerned with the taxonomy of a given group are most desirable if their classificational results are based on a comprehensive
phylogenetic hypothesis. If at all possible, a systematist should try to interpret his
taxonomic data with respect to the corresponding phylogeny rather than relying
on an intuitive character evaluation as a basis for a classification. Incidentally,
since the hierarchical structure of the organisms on earth is the result of a histori-
cally real, continuous sequence of species splittings, the hierarchy of organisms
as reconstructed by phylogenetic methods is in the end the only objective base for
a classification.

4.5 Zoological nomenclature: governing the
process of naming

Biologists are considered with a myriad items, and these items need to be named
to communicate concepts and hypothesis about these items in the biological
sciences. It is not only the millions of species that have to be named. The
hierarchical structure of the biodiversity of species and the myriad supraspecific
taxa result in an almost infinite number of items that in principle can be named.
Since names should be unequivocal in an international perspective, the formation
and practical handling of names of species and higher taxa obviously needs
internationally binding regulations. Nomenclature is the part of taxonomy that
regulates how names are assigned to taxa, and the underlying set of formal rules
of how nomenclature is applied in practical zoology is the International Code of
Zoological Nomenclature.

4.5.1 Ranking hierarchies: capacity and limits of
Linnaean categories

The Linnaean categories are terms that identify the hierarchical level in a classifi-
cation. Examples for categories are genus, family, and class. A limited number
of such categories date back to Carolus Linnaeus’ publications in the eighteenth
century, but the number of categories has increased continuously over time, the
better to reflect increasingly complex classifications. The rise of phylogenetic
methods in the last few decades has resulted in an increasing number of clado-
grams that depict complex, not directly linear relationships with a seemingly
infinite number of potential levels. It is, thus, not surprising that systematists face
a myriad practical problems when assigning Linnaean categories to classifications
based on cladograms resulting from phylogenetic analysis. Even in earlier phylo-
genetic publications, such as Hennig (1969), the conflicts between phylogenetic
results and categories were well known, and in that publication, Hennig replaced
the Linnaean categories by a system of consecutive numbers that directly reflects sister group relationships. However, his system received little attention from the beginning, because in less inclusive taxa, single numbers could be confusingly long (such as 2.2.2.2..4.6..1.1. Trichoptera, the sister group of 2.2.2.2..4.6..1.2. Lepidoptera). Hennig’s system seems to have disappeared from consideration.

Since that time, various nomenclatural systems have been proposed to eliminate or replace the Linnaean categories, one of the most recent and popular alternative concepts being the PhyloCode. It is beyond the scope of this chapter to present the characteristics, the advantages, and the failures of the PhyloCode, but I will focus on a critical assessment of the capacity and the limits of Linnaean system instead. It is still the most popular and widely accepted system of taxonomic ranking, and I will argue that this is not due to the uncritical persistence of antiquated principles but, conversely, for good reasons. The reader should consult one of the numerous papers on the Linnaean categories and their alternatives for more information, such as Cantino and de Queiroz (2002) in the World Wide Web, and Pleijel and Rouse (2003) in favor of the PhyloCode, and the rebuttals of arguments for the PhyloCode and its underlying theories by Nixon and Carpenter (2000, 2003), Carpenter (2003), Schuh (2003), and literature cited therein.

Ranking is an inherent property of any hierarchy in biological systematics, whether it is portrayed in a cladogram or by a formal Linnaean ranking scheme. Since ranking is included in any hierarchy, several unambiguous statements can be made concerning the structure of this hierarchy: for example, if taxon A is more inclusive than taxon B and actually includes B, then B does not contain A. However, the taxon names A and B themselves do not carry any reference per se about their relative position in the hierarchy, that is, whether A includes B, or vice versa, or whether A and B are of equal rank. In complex hierarchies, such as the hierarchy of organisms, the user is concerned with an overwhelming number of taxon names of different relative ranks, and he would be lost if any taxon bore an arbitrary name. Instead, communication is considerably more efficient if the ranking information is stated in addition to the taxon name. The Linnaean categories were invented exactly to convey this information by the use of standardized suffixes (for family-group names) and binominal nomenclature for species. As an example, the Cercopithecidae, the Old World monkeys, is usually ranked as a family (indicated by the suffix -idae) and, among others, contains the subfamily Colobinae (with the suffix -inae).

The Linnaean ranking system was, and still is, extremely successful, and this is partly because it is easy to understand and to learn. Alternative concepts that have been proposed to replace the Linnaean categories are explicitly rankless. Taxon names still exist, but they do not communicate any information at all
about their position in the hierarchy. The user can only understand the hierarchical structure by referring to the underlying cladogram. Hence, discarding Linnaean ranks and the binominal nomenclature results unnecessarily in the complete loss of important information on the taxa. This is not to say that the application of the Linnaean system is always unambiguous. Instead, there are numerous problems involved in taxonomic practice, particularly when portraying complex cladograms in a hierarchical classification. However, it seems unlikely that “[any] single system of nomenclature can ever possess all desirable attributes” (Schuh 2003 p 60).

4.5.1.1 Taxa and categories are not the same

There is a fundamental difference between a taxon and a category. As defined earlier, taxa are groups of real organisms that can be described and named. They can be monophyletic, nonmonophyletic, or phylogenetically untested. In contrast, categories are terms that can be assigned to taxa to connote a certain rank relative to other, more inclusive or less inclusive taxa. Taxa are the objects of biological systematics, whereas categories are a voluntary, notional tool that improves communication among systematists.

4.5.1.2 Categories and age

Linnean categories contain explicit and helpful information about the relative position of a taxon within a certain lineage. The central question in taxonomic practice is when we have a hierarchy of names that is supposed to reflect a hypothesis of relatedness (or at least overall similarity), how do we know which of the various levels should be ranked as an order or a family? To be honest, we actually never know, at least not precisely. The “alignment” of the hierarchy of taxa and the hierarchy of categories is determined by practical considerations, which is in the end a matter of subjectivity. However, there is a long lasting discussion of how to “objectify” the assignment of categories to a specific level in a classification. Actually, all efforts to do this have failed, and future efforts will probably fail. The reasons are easy to understand. Categories, as indicators of relative rank within a phylogenetic lineage of organisms, are inherently subjective, at least in a sense. It is a matter of usefulness for verbal communication and of convention and consistency, if, for example, taxon X is ranked as an order or a class. However, neither the hypothetically real hierarchy of organisms nor the categories themselves provide objective criteria how to applicate them.
However, several approaches to objectify categories have been published. Since phylogenetical reconstruction achieves to reconstruct historical events in the evolution of organisms, it is tempting to try to objectify the assignment of categories to taxa by using age as a criterion. This was already proposed by Hennig (1950, 1966), who discussed the problem of *absolute ranking* of higher taxa in exhaustive details. He suggested that the fossil record allows us to unambiguously define categories by the age of taxa to be classified. His example is the higher-level relationships of insects. Insects are generally considered to be ranked as a class, and since the oldest fossil insects, which belong in the Collembola, have been recorded from the Middle Devonian, the Collembola and their immediate relatives, which must have also been present at that time, would be assigned to what Hennig called a “class stage.” This resulted in the following ranking for Collembola: class Insecta, subclass Entognatha, infraclass Ellipura, microclass Collembola. Hennig, of course, realized that this approach is impracticable and arbitrary. He suggested to “reserve the well-known category designations (‘class’ and ‘subclass’) for the most important and morphologically isolated groups” (Hennig 1966 p 185), which can hardly serve an objective criterion. Hennig and subsequent workers were well aware that the age criterion of the categories across organisms or at least animals cannot work, simply because the major evolutionary radiations and subsequent diversifications took place in extremely different periods of time. As an example, the major lineages of the placental mammals that are generally ranked as orders arose in the Cretaceous (Murphy et al. 2001), and thus are comparable in age to the major lineages of Hymenoptera (bees, wasps, and ants), which are considered to be of family rank (Rasnitsyn 2002). If categories could be defined horizontally, the categorical hierarchies of Placentalia and Hymenoptera would have to be adjusted to each other, either by “downgrading” Placentalia to family rank or by upgrading Hymenoptera to class rank. This example clearly shows that a horizontal, *absolute* definition of the Linnaean categories is not only impractical but actually absurd.

However, a *vertical*, relative application of Linnaean categories along a phyletic lineage is a different matter and may be possible as a kind of evidence-based convention. An example is Goodman’s et al. (1998) cladistic analysis of the Primates based on DNA evidence and on extant and fossil morphological characters. Many traditionally recognized taxa appeared to be monophyletic, and their names and ranks as used in current literature were maintained. The ages of these clades were determined using dating evidence as provided by fossils and the model of local molecular clocks. Based on this time scale, clades of roughly equivalent age were assigned the same Linnaean rank. Goodman et al. (1998) analysis resulted in the following dating scheme (category/age [Mio years]/ geological period): semiorders/63/early Paleocene; suborders/58–50/late Paleocene to
early Eocene; infraorders/45–40/middle Eocene; superfamilies/39–29/middle Eocene to middle Oligocene; families/28–25/middle to late Oligocene; subfamilies/23–22/early Miocene; tribes/20–14/early to middle Eocene; subtribes/14–10/middle to late Miocene; genera/11–7/late Miocene; subgenera/6–4/late Miocene to early Pliocene.

It needs to be emphasized that an age-related Linnaean ranking might have merits only if applied to clades like the Primates with particular prerequisites. Primates have a widely accepted classificational framework (though the details are still disputed), quite a good fossil record, and cladistic analyses have been published with reliable taxon sampling. If the last common ancestor of the Primates is supposed to be as old as about 63 Mio years, and if the scientific community agrees to rank the Primate clade as an order, the highest level of the classification within Primates is defined: clades that arose in the early Paleocene are ranked as orders. Other, younger, fossils can be correlated with a particular rank accordingly. The fossil record, which might include gaps without fossil evidence, is supplemented by the model of the molecular clock, as proposed for Primates. This procedure allows applying the age criterion to any ranking level between the oldest reference point (the putative age of the last common ancestor of the Primates) and the youngest reference points (e.g., the latest splittings to clades generally ranked as subgenera, the youngest fossils).

Approaches like this pretend to be objective in correlating taxa and ranks, at least within the clade they were proposed for. However, it is easy to see that none of the criteria that define the age–rank correlation as proposed by Goodman et al. (1998) leads to an objectively settled ranking. It is a matter of convention if the clade named Primates, which is defined by a set of apomorphies and whose age of origin is defined by a particular fossil, is ranked as an order. Likewise, it is a matter of convention which of the clades within Primates, such as Cercopithecoidea, is assigned superfamily or any other rank. The same arbitrariness applies to the correlation of a clade and an age per se: with respect to the current state-of-the-art in Primate classification it might be useful to treat the putatively 18-Mio-years-old Cercopithecoidea as a superfamily, but other rank–age combinations might be similarly appropriate if suggested by the total hierarchy within Primates.

Nomenclatural stability is another important issue to be addressed to any classification concept. Stability of nomenclature should minimize changes in the names as they are adopted to changing scientific concepts. Age-defined ranks are, at any hierarchical level, extremely sensitive to the accuracy of the underlying dating of the latest common ancestors. New dating evidence, such as newly discovered fossils or new analytical methods in paleontology or molecular biology, may suggest fundamentally different dates of the origin of particular groups.
If, for example, the origin of Primates dates back to 80 Mio years as suggested by Tavaré et al. (2002), the age-related ranking scheme of Goodman et al. (1998) would have to be adopted to the new dating by stretching the ranking range to fit to the longer period of time or by adding more category levels.

The standard approach to calibrating a classification for an age-related application of ranks is to use the earliest known fossils of a particular taxon and to equate their age with the time of origin of this taxon. The resulting age of the taxon under discussion is, obviously, hypothetical and only gives us the so-called “terminus post quem non,” that is, the point of time after which the clade cannot have originated. In other words, the fossil record only settles the minimum age of the group. This can potentially lead to a serious underestimate of the true time of origin of a clade, and hence, of all other dating results inferred secondarily. As has been pointed out by Martin (2002), this underestimation increases if the fossil record is very patchy. Since this might be the case in many organisms (Tavaré et al. 2002), the oldest fossil of a given clade might be considerably younger than its true stem species. As a result, age-defined ranking schemes are considerably different depending on whether the dating as inferred directly from the fossil evidence, or the dating as indirectly estimated by molecular clock models or estimations of extinction rates and the reliability of the fossil record, is considered.

These applications and problems show that an age-defined, objective correlation of Linnaean categories and clades is as unrealistic, if not impossible, as any other effort toward an absolute definition of ranks. In some cases, such as Primates, age-defined ranks may be useful to some extent, although it must be emphasized that the correlation of ranks and clades even in such cases is at most intersubjective, being based on a set of conventions of the respective scientific community. Changing conventions, changing dating evidence, and other factors may considerably alter the putatively objective ranking of Primates. The Linnaean categories are a notional representation of the hierarchical structure of monophyletic taxa representing an inferred, hypothetical sequence of speciation events (or, if a cladistic analysis is still lacking, an a priori assumption of relatedness inferred from overall similarity). As a consequence, the Linnaean system is inherently relative, that is, Linnaean ranks denote the hierarchical position of a given taxon relative to other members of the same clade. As has been paraphrased by Schuh (2003 p 60), “[t]he primary strength of the Linnaean system is its ability to portray hierarchical relationships.” It must, however, be emphasized that Linnaean categories are inherently subjective in that the exact correlation of a taxon and its rank are not determined by the taxon or the category themselves.
4.5.1.3 Ranking fossils

Fossil and recent taxa are basically treated in an identical way in taxonomy. However, if fossils are included in a cladogram, a number of practical terminological problems arise. Fossils are part of the stem lineage of a given taxon, and each single fossil taxon is considered the sister group of the next, less inclusive, taxon. If sister taxa are given the same rank in the hierarchy, each single fossil species that is considered to be the sister species of a taxon that is ranked as an order must be assigned order rank as well. If several fossil species of a given stem lineage in a continuous sequence of sister group relationships with the next less inclusive taxon are known, each of these single species would have to be given a high rank, which would, for example, result in a proliferation of taxa ranked as orders along a single lineage. In principle, each of the pairs comprising a fossil species and its higher sister taxon could also be assigned a new name.

As a consequence, giving a high rank to each of the fossil species would not only appear to be overblown, but this would also ignore the fact that the exact sequence of fossils along a stem lineage cannot be determined accurately in most cases. As an alternative, Patterson and Rosen (1977) have suggested to order fossils according to their hypothetical phylogenetic position but to leave them unranked. Instead, each of these fossils is named *plesion*. The term *plesion* replaces any rank above the genus level and to tentatively place fossils in cladograms without affecting the hierarchical structure of names and their ranks.

4.5.1.4 Linnaean categories as a communication tool

Systematists should be aware that the Linnaean categories are a tool for communication. As Griffiths (1976 p 168) stated, there is no logical reason why taxa *must* be ordered into categories. Linnean categories do not inherently imply information about the age of a taxon (except if employed within a certain clade by convention, as e.g., in Primates), the size of a taxon (in terms of species numbers), genetic distinctiveness, or phylogenetic relatedness. The application of a particular rank to a particular taxon is subjective but must fit in the overall hierarchy of the more inclusive taxa.

But Linnean categories directly mirror the complex, hierarchical relationships in a given clade. Along this lineage, the nestedly ordered categories notionally reflect the nested order of organisms and explicate the relative position of a given taxon to other taxa of the same lineage. The demand for an abolition of the Linnean categories without substitution involves well-known practical problems and would result in the loss of practical and useful information. There is a
myriad practical problems involved in the implementation of the Linnaean categories, but this is also the case in any other notional system of ranking. “No single system of nomenclature can ever possess all desirable attributes: i.e., convey information on hierarchical relationships, provide ... stability in the names . . ., and provide simplicity and continuity in communicating the identities of the taxa and their relationships” (Schuh 2003 p 60). Since the discussion about the Linnaean categories reflects the problems how best to express scientific contents linguistically, applicability is the central criterion to assess the capability of any ranking concept. The search for the “best” biological nomenclature corresponds to the search for compromise between theoretical demands and practical necessities.

4.5.2 The International Code of Zoological Nomenclature

Systematists and, secondly, subsequent users of the results of biological systematics, deal with a myriad objects: each taxon, if considered a species or ranked as a genus, family, or order, can be recognized as a biological entity, which, to a varying degree, is distinctly different from other such objects. To enable communication about this objects, in principle requires the application of unique names, which unequivocally refer to the taxa. Even in pre-Linnaean times, the need to point to groups of organisms by names was obvious. However, before Linnaeus, the names given to taxa were thought to be descriptive or diagnostic terms or phrases rather than proper names, as we see taxon names today. Although easily recognizable organisms might have been given single names (uninomina) or, if two closely related forms were known, two-word (binominal) names were coined. Increasing knowledge of the diversity of forms made complex, descriptive labels necessary to meet the criterion of uniqueness (e.g., the gentian species that is today known as *Gentiana ciliata* was named *Gentiana angustifolia autumnalis, minor floribus ad latera pilosis* by Bauhin (1623) (Mägdefrau 1992); note that Bauhin and other pre-Linnaean authors already employed unique genus names as more inclusive taxa). Linnaeus, however, introduced a strictly formal nomenclatural system, based on a unique binominal species name. The increasing exploration of remote geographic regions made scientists realize that they are concerned with a seemingly infinite organismic diversity. Due to new discoveries and new techniques taxon names proliferated. However, it was not only the increase in species descriptions that lead to a rapid increase in the total number of taxon names. Inconsistent or even contradictory terminological systems, development of their own rules by many scientists, replacement of already published names that were considered to be inappropriate or incorrectly
formed, and other factors, resulted in the strong need for general rules of nomenclature.

Beginning with Linnaeus’ works in the eighteenth century, an elaborate body of conventional regulations was developed to provide such general rules. The valid rules of nomenclature in zoology are contained in a judicial text called the International Code of Zoological Nomenclature, here abbreviated as the Code, whose current edition is the fourth. The Code provides a complex set of rules and aims to cover as much of the complexity of the recent and historical naming process as possible. However, cases might show up in which the strict application of the Code might actually threaten nomenclatural stability, and in such cases every scientist is free to apply to the International Commission of Zoological Nomenclature to set aside priority or other provisions of the Code to increase stability. Many provisions of the Code are the result of compromise between conflicting principles, which partly date back to the very beginning of biological nomenclature. Hence, the rules fixed by the Code are far from written in stone but are under constant development. As has been emphatically pinpointed by Mayr and Ashlock (1991 p 386), “all good law is living law.”

An exhaustive description of the Code is far beyond the scope of this chapter, but a few basic principles will be considered here.

4.5.2.1 Freedom of taxonomic thought

The Code does not tell anybody how to classify something or how to recognize species or other taxa. These are scientific decisions that are totally under the responsibility of the systematist. The Code provides rules for a system of communication, a “language,” for communicating such taxonomic decisions. This “freedom of scientific thought” is expressly guaranteed in the Preamble of the Code, and, indirectly, emphasizes the role of nomenclature in biology as a notional process acting subsequent to the scientific process. As an example, the Code permits a scientist to continue using a name that is considered to be a junior synonym by another author. Conversely, the freedom of taxonomic thought includes also the freedom to make taxonomic mistakes that, as a consequence, do not affect nomenclature. If, for example, a name is erroneously placed in synonymy, its validity is not affected.

4.5.2.2 New names and old names

The process of naming a newly discovered species properly, that is, in accord with the Code, is surprisingly simple. The Code explicitly provides a set of provisions
to be met when describing a species, some of which appear to be commonsense principles. To be usable or, to use the proper term, to be available, a name must, first, be published. The Codes clearly indicate that distribution “in numerous identical copies . . . for permanent record” constitutes a publication. The use of the 26 letters of the Latin alphabet for the taxon name is mandatory, although the descriptive text may also be written using a different alphabet. A name must, of course, in principle be unique but on different levels: a species name must only be unique in its own genus (there are several animal species with popular species names like viridis, meaning green, but no two species (or subspecies) in a genus may bear this name), whereas a genus name must be unique among all animals. However, it is wise to avoid species names that are already in use in closely related genera. In the future, genera might be combined due to changing evidence, which would result in homonymy of such species names. A new species name “must be explicitly indicated as intentionally new,” that is, by adding a Latin term like “new species” to the newly proposed name or an equivalent phrase or abbreviation. The Code also requires the consistent application of binominal nomenclature, which does not apply to names of taxa at ranks above the family group). A new species description must also include an explicit fixation of the name-bearing type specimen(s). A type is an individual specimen to which a given name is attached. In cases in which a taxon previously considered a single species is later split into two species, the species to which the type specimen belongs retains the previously given name. This type method is important in taxonomy and will be presented in more detail later. Finally, the Code provides a set of mandatory regulations of more linguistic nature, how new names are to be formed. The most important provision is that names must be either Latin or latinized or they must be so constructed that they can be treated as Latin words. Further rules concern practical details such as how names are formed from personal names and other aspects of the correct spelling.

Every scientist who plans to propose a new name for a taxon should consider these simple provisions to be sure that the new name complies with the Code. However, the greater part of the Code is devoted to less simple cases, and most of the problems are caused by names proposed in the past. A smaller part of difficulties systematists face when dealing with previously published names might be caused by the erroneous application of the Code or by erroneous taxonomic decisions. The majority of problems, however, simply portray the continuously changing concepts and applications in the 250-year-old history of biological systematics. Such changes might affect scientific content (e.g., by the invention of alternative species concepts) or might be more formal (e.g., the explicit designation of a name-bearing type was not mandatory at the beginning of biological systematics). As a result, historical names are not only hard to locate
in the literature in many cases, they often require time-consuming, elaborate efforts to clarify the correct original spelling, the presence of type material, the identity of the species, the correct date of publication of the original description, and several other potential difficulties. Two of the most common problems that come up when dealing with historical names are synonymy and homonymy.

The fact that a considerable portion of taxonomic practice is unusually time consuming when compared to the outcome, i.e., the clarification of a single species name, is part of the legacy of about 250 years of biological systematics. In many groups of organisms, taxonomists are condemned to spend most of their time interpreting the work of pre-twentieth-century systematists. “The past [often] acts as a dead weight on the subject, the complex synonymy and scattered type material deterring anyone from attempting a modern revision” (Godfray 2002 p 17).

4.5.2.3 Validity, synonymy, and homonymy

As I have explained earlier, a taxon name must fulfill a set of conditions to be available in terms of Code compliance. Among the available names, only the oldest available name is what systematists call valid and has priority over the younger names. The valid name of a taxon is sometimes referred to as the correct name, which means that open nomenclatural problems have been solved and that a single name retained, that under the provisions of the Code, is the only approved one. The date of publication of the original description of a name is thus of crucial importance and must be carefully determined. It must be emphasized that the principle of priority applies to the species, genus, and family levels only but not to levels above the family rank.

In by far the most cases the valid name is the oldest name. Exceptions occur when the oldest name, in case of a species, is already in use in the genus, resulting in homonymy. Then the younger of the homonymous names would no longer be valid but must be replaced by the next available name of the same species, if any exists. An example is a subspecies of the common Chimpanzee, Pan troglodytes verus Schwarz, 1934 (originally described as a subspecies of Pan satyrus Linnaeus, 1758, the species name of which was suppressed by the International Commission on Zoological Nomenclature in 1999). The oldest available name of this subspecies is Simia chimpanse Matschie, 1904, but, despite having priority due to the early publication date, Matschie’s name cannot be used because it is a junior homonym of another species, Satyrus chimpanse Mayer, 1856. Since Satyrus chimpanse Mayer is considered to be identical to, that is, a synonym of, the
nominate subspecies, *Pan troglodytes troglodytes*, the species group name *chimp-\textit{anse* occurs twice within *Pan*, by Matschie, 1904, and by Mayer, 1856, respectively. This is called *secondary homonymy*, since the two names under discussion were first proposed in different genera and got in conflict after being considered to belong in the same genus. *Simia chimpanse* Matschie was published half a century later than *Satyrus chimpanse* Mayer, and thus as the younger of the two names must be replaced. The next available name is *Pan satyrus verus* Schwarz, published in 1934, which then becomes valid. In many species, however, no name is available other than the originally proposed name. Then the revising author has the authority to propose a replacement name, which then becomes available with the reviser as the correct author and the date of his publication.

If a single taxon is given two or more names, each of these names is a *synonym*. The earliest published synonym is referred to as the *senior synonym*, even if it is considered to be the valid name. Any other earlier names of the same taxon are called *junior synonyms*. In most cases in practical taxonomy, systematists are confronted with the question if two independently published species names that were based on different name-bearing types actually represent the same species. This kind of synonymy with different types is called *subjective synonymy* because it expresses the scientific conviction of the revising author that the two names refer to the same thing. Names applied to a species on the basis of the same type specimens are *objective synonyms*.

### 4.5.2.4 Typification: establishing objective reference points

The type method is a guiding principle in biological systematics. In taxonomic practice the identity of species, particularly when described in the past, is often difficult to establish due to insufficient descriptions or illustrations. As a result, historical species descriptions might apply to more than one currently recognized species. Due to new evidence, taxa might be split up into two or more less inclusive taxa. In all such cases, it might be difficult to establish which of the newly separated taxa has to retain the original name and which should be newly named. This problem is solved by the invention of objective reference points, the types. Types in the broadest sense are zoological objects to which a name is firmly tied. The type of a species is always a single specimen, the type of a taxon at genus rank is a species, and finally, the type of a family-rank taxon is a genus. If a previously described species is considered to be actually composed of more than one species, the type specimen fixes the original name to the species, to which it belongs. The same holds for taxa at the genus and family ranks.
The Code permits the use of a few different kinds of types in taxonomic descriptions. At the genus and the family level only type species and type genera exist, respectively. A **holotype** is always a single specimen, which is either the only specimen available for the original description, or it is explicitly selected out of the original series of more than one specimen to give it precedence over the other specimens, which then are called **paratypes**. In current species descriptions, the explicit designation of a holotype is mandatory. Paratypes do not have special standing under the Code but should be explicitly mentioned as members of the original type series by calling them paratypes. If the original describer in older publications mentions more than one specimen but did not select a holotype, all type specimens are the **syntypes**. A syntype series is of little help in conflicting taxonomic situations because they might actually belong to more than one species. In such cases, a revising author has the authority to subsequently select a single specimen from the original type series, which then would be called a **lectotype**. Accordingly, the remaining syntypes would be the **paralectotypes**, which, again, are only of secondary importance. If the type material of a species is considered to be lost, the revising author may select a specimen that was not part of the original type series, as a **neotype**. In some cases, the lack of type material is of minor importance because the identity of the species is obvious. An example is *Homo sapiens* Linnaeus, 1758, for which no type specimen or specimens have ever been settled (Spamer 1999). A neotype must be designated only if this action facilitates nomenclatural stability so that the name is properly and consistently applied. Since a neotype is required in cases with unsettled and problematic nomenclature, one can almost never be completely sure if the neotype and the original type are really conspecific. Thus, designation of a neotype might be risky and should only be done if absolutely necessary. Numerous other kinds of types have been used in the literature, but only those above are still permitted under the current Code. Examples of types that are not regulated by the Code are allotypes (a specimen of the opposite sex of the holotype, actually simply a paratype) or toptype (a specimen originating from the type locality of the species to which it is thought to belong).

Incidentally, it is a widely held belief that the type specimen or specimens of a newly described species must be dead or should be killed sooner or later to be finally deposited in a museum collection or in another kind of persisting and accessible environment. This means that in most cases a new species description is based on one or more specimens that are physically available to the scientist during the process of writing the description. The Code requires that the specimen (the holotype) or specimens (syntypes) on which the name is based must be explicitly stated and accompanied by description or diagnosis to differentiate the new taxon. However, this requirement does not include any mandatory
provisions about the physical availability, detailed morphological description, or depository of the type specimen (Wakeham-Dawson et al. 2002).

A type specimen need not be especially typical for the species to which it belongs. Its function is simply that of a “name bearer” (Simpson 1961). There are several potential criteria for how to select a holotype out of the entire material that is available to a scientist. If the species is differentiated from its congeners by means of complex morphological features of the male genitalia (as is common in invertebrates), it might be appropriate to make a male the holotype. Sometimes not all type specimens are complete or show the diagnostic characters equally well. It would make sense to choose as the holotype a specimen, which exhibits most or all of the characters that are of diagnostic value at the time of description and with respect to new future discoveries as far as can be predicted.

The description of a new species is always based on all specimens available at the time of description. Since the holotype is just a name bearer, its function is not to serve as the only basis of the original description (except when only a single type specimen is known).

### 4.5.2.5 Fossils, fragments, and heterogeneous type specimens

The Code applies to both living and extinct animals, as is explicitly stated in its Article 1. Thus, the principles of the type method as briefly presented earlier with reference to recent species are equally applicable in paleontology. However, many fossils, particularly complex multistructured organisms, such as vertebrates, inherently lead to practical problems. In most fossil vertebrates, only hard structures are fossilized and thus preserved. In the course of diagenetic processes, soft tissues that connected the hard elements in the living animal are usually decayed so that the resulting skeleton falls apart. In the ideal case, the preserved elements remain in their original position after embedding in a fossilization matrix so that the elements can easily associated with each other. In many cases, subsequent physical processes might have altered the relative position of the elements or have destroyed parts of the organism.

Fossil hominids are never completely preserved, not even with respect to bones. Sometimes just a single preserved bone is sufficient to provide evidence for taxonomic identity. If a single bone is the basis of a new taxon, just the bone is referred to as the holotype. An example is the holotype of *Homo heidelbergensis* Schoetensack, 1908, which consists of a complete, adult mandible. There is no ambiguity about the formal basis on which the new species is based. If more than a single element is available, association to individual can be difficult if the elements are not in the original position relative to each other but are more or
less dislocated or destroyed. An example is the composition of the types of *Homo habilis* Leakey, Tobias, and Napier, 1964. Several fragments assignable to the new species were available such as mandibles, isolated teeth, parietals and hand bones, and cranial fragments. Due to external evidence, such as the position of the fragments as found during excavation, the fragments could be associated with five individuals. The holotype of *H. habilis* consists of a mandible with dentition and the associated upper molar, parietals and hand bones, originating from a single juvenile individual.

These examples show another principle of the type method. A type is always a zoological object but is not necessarily something that is or is close to a complete organism. A single tooth, an isolated wing, or any other part of an organism is in principle sufficient to serve as the basis for a new species. In a way incompleteness is an inherent attribute of most preserved specimens because many preservation techniques result in the loss of some characters. Pinned, that is, dried insects and stuffed, mounted vertebrates have usually lost all or most of their soft tissue. Thus, even if a specimen has lost almost the entire physical structure except for a single tooth, this incompleteness of a specimen does not prevent potential type status. It is, however, wise to associate isolated fragments such as a molar and a mandible to a single type specimen only if the evidence is unambiguous. If subsequent evidence is found that the fragments actually belong to different individuals, this might cause a lot of unnecessary nomenclatural confusion.

References

Bauhin K (1623) Pinax theatri botanici. Basel
Groves C (2001a) Why taxonomic stability is a bad idea, or why are there so few species of primates (or are there?). Evol Anthropol 10: 192–198


5 Quantitative Approaches to Phylogenetics

Kaila E. Folinsbee · David C. Evans · Jörg Fröbisch · Linda A. Tsuji · Daniel R. Brooks

Abstract

We review Hennigian, maximum likelihood, and different Bayesian approaches to quantitative phylogenetic analysis and discuss their strengths and weaknesses. We also discuss various protocols for assessing the relative robustness of one’s results. Hennigian approaches are justified by the Darwinian concepts of phylogenetic conservatism and the cohesion of homologies, embodied in Hennig’s Auxiliary Principle, and applied using outgroup comparisons. They use parsimony as an epistemological tool. Maximum likelihood and Bayesian likelihood approaches are based on an ontological use of parsimony, choosing the simplest model possible to explain the data. All methods identify the same core of unambiguous data in any given data set, producing highly similar results. Disagreements most often stem from insufficient numbers of unambiguous characters in one or more of the data types. Appeals to Popperian philosophy cannot justify any kind of phylogenetic analysis, because they argue from effect to cause rather than cause to effect. Nor can any approach be justified by statistical consistency, because all may be consistent or inconsistent depending on the data being analyzed. If analyses based on different types of data or using different methods of phylogeny reconstruction, or some combination of both, do not produce the same results, more data are needed.

5.1 Introduction

Formalized by Willi Hennig in 1950, phylogenetic systematics has emerged as a universal and transparent method for generating and evaluating phylogenetic hypotheses. Over the last 50 years, it has developed into a research program ostensibly embraced by a majority of evolutionary biologists who are interested in exploring the patterns and processes of evolution. However, this apparent unity is misleading. As researchers from diverse fields began putting the theory and methods into practice, they approached their studies from different perspectives.
and using different types of data. This has generated a multitude of quantitative methods and research strategies, the efficiency and validity of which are fiercely debated in the literature.

These debates stem partly, perhaps largely, from the nature of phylogeny reconstruction. Unlike much of physics and chemistry (astronomy and astrophysics being notable exceptions), in which experiments in a hypothetico-deductive framework aim to be predictive with respect to spatiotemporally invariant laws, the reconstruction of phylogenies deals with a singular history and thus is descriptive and retrodictive. Attempts to infer causal processes for evolutionary patterns must be based not only on evolutionary patterns of relationships but also on corroboration from independent data (Brooks and McLennan 2002). Despite these obstacles, evolutionary biologists strive to provide as accurate an approximation of evolutionary history as possible.

Paleoanthropologists, like all paleontologists, should be especially concerned with the accurate reconstruction of primate evolution, partially because fossils provide some of the most powerful evidence supporting evolution, the unifying theory of biology. More particularly, paleoanthropologists are the curators of information about the most fascinating evolutionary story of all, the story of Us. As a critical part of evolutionary biology, paleoanthropology thus stands to benefit enormously from participation in phylogenetic research programs. Collaborating with researchers in diverse fields and exploring new methods will allow paleoanthropologists to refine their hypotheses and methods while placing them within the larger context of biotic evolution on this planet.

5.2 Fount of stability and confusion: A synopsis of parsimony in systematics

The principle of parsimony (Latin parcere, to spare) is also known as the principle of simplicity. The principle is often connected to the English philosopher and Franciscan monk William of Ockham (ca. 1285–1349), who advocated the use of the principle so forcefully that it is also known as “Ockham’s razor”: Pluralitas non est ponenda sine necessitate (plurality should not be posited without necessity) and non sunt multiplicanda entia praeter necessitatem (entities should not be multiplied unnecessarily). In this sense, the principle represents an epistemological tool or rule of thumb, which obliges us to favor theories or hypotheses that make the fewest unwarranted, or ad hoc, assumptions about the data from which they are derived. This does not necessarily imply that nature itself is parsimonious. Aristotle (350 BCE) articulated an ontological basis for the principle of parsimony, the postulate that “nature operates in the shortest way possible”
and “the more limited, if adequate, is always preferable” (Charlesworth 1956). This sense of the principle postulates that nature is itself parsimonious in some manner. Phylogeneticists have used the term “parsimony” in both senses, resulting in much confusion and, from our perspective, unnecessary conflict.

The most important concept introduced by Hennig (1950, 1966) was the stipulation that we should assume homology in the absence of contradictory evidence. Although now known as Hennig’s Auxiliary Principle, this concept lies at the foundation of evolutionary theory. “[p]erhaps the correct way of viewing the whole subject, would be, to look at the inheritance of every character whatever as the rule, and noninheritance as the anomaly” (Darwin 1859 p 13), and “Mr. Waterhouse has remarked that, when a member belonging to one group of animals exhibits an affinity to a quite distinct group, this affinity in most cases is general and not special” (Darwin 1872 p 409). Hennig’s argumentation method is clearly intended to maximize hypotheses of homology and minimize hypotheses of homoplasy, which invokes the principle of parsimony by avoiding the assumption of unnecessary ad hoc hypotheses of parallelism. In the Hennigian system, if evolution were parsimonious, all traits would be logically consistent with the true phylogeny—there would be no conflicting relationships suggested by any set of traits, that is, there would be no homoplasy. The Auxiliary Principle implies that there will be conflicts in the data, which should be resolved in favor of the hypothesis postulating the fewest number of assumptions of multiple origins (homoplasy) over single origins (homology). Contemporary Hennigians assert that both the Auxiliary Principle and the use of parsimony are logical requirements of any attempt to reconstruct phylogeny; if one were to assert that all similarities were due to homoplasy, there would be no evidence of common descent, and thus no evidence of evolution. Therefore, if one is going to study evolution, one must use a method that is capable of finding evidence of evolution. Likewise, if one is going to invoke the Auxiliary Principle, one must invoke it for all traits, thereby choosing the phylogenetic hypothesis that minimizes the total number of violations of the Auxiliary Principle for a given set of data. In this manner, the Auxiliary Principle is an epistemological tool practically synonymous with the principle of parsimony (Farris 1983; Wiley et al. 1991). Wiley (1981) suggested four main assumption of phylogenetics: (1) evolution occurs and has occurred, documented by the characters of different species; (2) each species is a historically unique mosaic of plesiomorphic, synapomorphic, and autapomorphic traits; (3) before the analysis we do not have knowledge about which characters are homologous and homoplasious; and (4) we do not know beforehand what the phylogenetic relationships are, nor do we know the relative or absolute rates of divergence. The presumption of homology embodied in Hennig’s Auxiliary Principle is not an a priori assumption in the sense of a formal
model, because the method is designed in part to recognize all mistakenly presumed homologies as homoplasies.

Edwards and Cavalli-Sforza (1963, 1964) reconstructed a tree of extant human populations based on frequencies of blood-group alleles, using an approach they developed and called the “Method of Minimum Evolution.” Their studies originally aimed to present a maximum likelihood method for phylogeny reconstruction, but their algorithm for a likelihood approach did not work. Edwards (1996 p 83) later emphasized that “[t]he idea of the method of minimum evolution arose solely from a desire to approximate the maximum likelihood solution,” that is, from a maximum likelihood model based on the assumption that evolution has been parsimonious. Felsenstein (2004 p 127) characterized the method of minimum evolution as a parsimony method, while at the same time not seeing a direct connection between Hennig’s Auxiliary Principle and the principle of parsimony, e.g., “[i]t is not obvious how to get from this ‘auxiliary principle’ to the parsimony criterion” (Felsenstein 2004 p 138). This reveals that for Felsenstein and like-minded phylogeneticists, parsimony is an ontological issue, whereas Hennigians see it as an epistemological issue.

There are two critical distinctions between these positions. The ontological perspective on parsimony requires first that evolution be parsimonious in some manner, usually as defined by certain assumptions and parameters of a model; and second, that the resulting phylogenetic hypothesis be accepted as true so long as the model is accepted as true. Practitioners are thus preoccupied with the accuracy of their results, and believe it is possible to develop means by which their preferred hypotheses can be verified with respect to the true phylogeny. The Hennigian or epistemological use of parsimony does not imply that the evolutionary process itself is parsimonious. In fact, it suggests that evolution has been so complex that we should always expect to find conflicts in the data, which will require the use of a logical decision-making principle to resolve. An important corollary of this perspective is that there need be no necessary connection between the most parsimonious hypothesis and truth. Practitioners are thus preoccupied with the empirical robustness of their results. The expectation is that if the most parsimonious hypothesis is not true, the accumulation of additional data will force phylogeneticists to abandon it in favor of a new most parsimonious hypothesis; they do not believe that their hypotheses can be verified, but do believe that they can use new data to falsify all or parts of previous hypotheses. Phylogeny reconstruction is thus an open-ended process involving a potentially endless search for information. If, at some point in the future, the accumulation of data has led to a situation in which the phylogenetic hypothesis for a given group is no longer changing with the addition of new data, Hennigians
may express the belief that the hypothesis has approached the truth as closely as possible, but in principle it is never appropriate for a Hennigian to claim to have the true phylogeny. Hennigians do feel justified in claiming that they have the most robust hypothesis possible for any set of data.

Today numerous quantitative methods for reconstructing phylogenetic trees are applied to multiple kinds of characters. These methods can be divided into two main types, commonly called parsimony (invoking epistemological parsimony) and likelihood (invoking ontological parsimony) approaches.

5.3 Epistemological parsimony: The Wagner algorithm

In September 1965 two articles on phylogeny and parsimony appeared. Wilson (1965) introduced a “consistency test for phylogenies based on contemporaneous species.” His null hypothesis was that all characters that are used for a phylogenetic analysis are unique and unreversed. In order to pass Wilson’s consistency test, the taxa defined by these characters must be nested and these nested conditions must persist as new species are added to the tree. Colless (1966) was concerned that more than one cladogram might pass the consistency test, that a polyphyletic character state might mistakenly be regarded as unique and unreversed, and that the taxa are in the first place grouped on the basis of similarities. Wilson (1967 p 104) asserted that his consistency test was internally sound, but that he shared one of Colless’ main concerns, which “is the lack of efficient methods for selecting the character states.”

That concern was discussed in the second article published in 1965 in which Camin and Sokal (1965) presented the first algorithm for applying the parsimony criterion to phylogenetics and first applied the term “parsimony” to a method of phylogenetic inference. They used a group of imaginary animals (Caminalcules) possessing a number of morphological characters that could change according to particular rules. Thus, the “true phylogenetic tree” was known and could be compared to trees that were achieved by different methodologies. Camin and Sokal (1965) found that the trees that most closely resembled the “true phylogeny” required the least number of changes in the morphological characters, which seems to invoke an epistemological use of parsimony. However, they claimed that their technique examined “the possibility of reconstructing cladistics by the principle of evolutionary parsimony” (p 312), saying that “the correctness of our approach depends on the assumption that nature is, indeed, parsimonious” (pp 323–324), an appeal to ontological parsimony. Significantly, Camin and Sokal produced a computer program implementing their method, demonstrating for the first time that quantitative phylogenetic analysis could be performed on as
objective a basis as phenetics, thereby undermining one of the strongest arguments in favor of phenetics over evolutionary approaches to systematics (Sokal and Sneath 1963). Their algorithm was unwieldy and inefficient for larger data sets, and was never fully adopted nor effectively programmed and made available for widespread use.

Soon afterward, Kluge and Farris (1969; also Farris 1970) presented a new algorithm for reconstructing phylogenetic trees as well as searching among several trees for the most parsimonious tree for a given data set. They named their method “Wagner parsimony” in honor of W.H. Wagner, who formalized an older approach (Mitchell 1901, 1905; Tillyard 1921; Sporne 1949, 1953; Danser 1950) called the groundplan-divergence method (Wagner 1952, 1961, 1969, 1980), which formed the basis for Kluge and Farris’ algorithm. Kluge and Farris (1969) also discussed explicitly their perspective that the use of the parsimony criterion did not assume that evolution itself is parsimonious, clearly invoking an epistemological use of the principle.

5.3.1 The Wagner algorithm

Kluge and Farris’ (1969) method minimizes the Manhattan distance between members of a set of taxa via the creation of hypothetical taxonomic units. The first Wagner algorithm (in later papers termed the “simple Wagner algorithm”) is constructed as follows:

**Definition:** \( X(A, i); \) state = \( X_i \), of character \( i \), for Taxon A

\[ D(A, B) = \sum_i [X(A, i) - X(B, i)] \] (1)

\( D(A, INT(B)) \) is the difference between A and interval B;

\[ D(A, INT(B)) = \frac{D(A, B) + D(A, ANC(B)) - D(B, ANC(B))}{2} \] (2)

Step 1. Choose an ancestral taxon.
Step 2. Find the operational taxonomic unit (OTU) that is the least distance from the ancestor, using Eq. 1. Connect the taxon to the ancestor, forming an interval.
Step 3. Find the next taxon that is next least distant from the ancestor.
Step 4. Find the interval from which the taxon found in step 3 differs least, using Eq. 2.
Step 5. Attach the taxon found in step 3 to the interval found in step 4 by constructing an intermediate (hypothetical ancestor), Y, and insert it into the
tree at this interval. The character states of Y will be the median of the three nodes surrounding this newly created intermediate.

Step 6. If any taxa remain unplaced, go to step 3, otherwise stop.

Example:

Using the same four-taxon, five-character statement as was used in the groundplan-divergence example (Table 5.1)

Table 5.1
Exemplar of a four-taxon, five-character statement

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>K</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>L</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>M</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>N</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Step 1. As in the example for groundplan-divergence, K is determined to be the ancestor.

Step 2. Requires us to find the difference between the ancestor (or outgroup) and each of the other taxa using equation 1.

\[
D(L, K) = \sum_i [X(L, i) - X(K, i)] \\
= [1 - 0] + [0 - 0] + [0 + 0] + [0 + 0] + [0 + 0] \\
= 1
\]

So the difference between L and ancestor K is 1. We must then determine the difference between the ancestor (K), and each of the other taxa in the matrix, using the same equation.

\[
D(M, K) = 3 \\
D(N, K) = 4
\]

We find that the difference between L and K is the least, so we connect L to K via an internode, as shown below. This interval will be known as interval L (INT (L)) (Figure 5.1):

Step 3. The next step is to determine which taxon to add to the cladogram next. It is the one that is the next closest (least different) from ancestor K. In this case, it happens to be taxon M (with a difference of 3 from above).

Step 4. To figure out the interval to which we can add M, we must calculate the distance between M and each of the intervals on the diagram. At this point, there
is only one interval so it is self-evident to which interval the taxon will be added, but the calculation is:

\[
D(M, \text{INT}(L)) = \frac{D(M, L) + D(M, \text{ANC}(L)) - D(L, \text{ANC}(L))}{2}
\]

\[
= \frac{2 + 3 - 1}{2}
\]

\[
= 2
\]

Step 5. Taxon M is then added to INT(L) via a hypothetical intermediate, Y (Figure 5.2).

Hypothetical intermediate Y is given the character states that are the medians of those taxa to which it is connected (L, M, and K). It is then added to the character matrix as in Table 5.2:

Step 6. The next taxon to be added to the diagram is the last remaining one, N. We now must determine the interval to which we must add N. To do this, we find the difference between N and each of the intervals. INT(M), INT(L), and INT(Y) using equation 2.
\[ D(N, \text{INT}(M)) = 2 \]
\[ D(N, \text{INT}(L)) = 2.5 \]
\[ D(N, \text{INT}(Y)) = 3 \]

N is found to differ least from \text{INT}(M), so it is added to that interval via a hypothetical intermediate, X, whose character states are the median of M, N, and Y. The network is now complete (Figure 5.3).

Farris (1970) concluded that it was unnecessary to have an ancestor from which to begin the construction of the tree. He observed that the choice of ancestor of a given group of taxa changed the topology of the tree. Since the “simple” algorithm did not impose directionality to the evolution of the group, he reasoned that the choice of ancestor is not crucial. Since parsimony assumes the least about the way evolution works, then choosing one taxon as an ancestor would be assumption about the status of that taxon. He thus argued that a rootless network would reduce the dependency of the form of the tree on the

\begin{table}[h]
\begin{tabular}{|c|c|c|c|c|c|}
\hline
 & 1 & 2 & 3 & 4 & 5 \\
\hline
K & 0 & 0 & 0 & 0 & 0 \\
Y & 1 & 0 & 0 & 0 & 0 \\
L & 1 & 0 & 0 & 0 & 0 \\
M & 1 & 1 & 0 & 0 & 1 \\
N & 1 & 1 & 1 & 1 & 0 \\
\hline
\end{tabular}
\end{table}

Table 5.2
New character matrix including hypothetical ancestor Y

Figure 5.3
Cladogram of relationships between taxa in Table 5.2
ancestor. For the creation of networks, he used a method for creating networks that minimized the length of the intervals between taxa (symbolized by nodes), using the shortest network connections method of Prim (1957; Sokal and Sneath 1963). Farris differentiated his use of this method from previous phenetic applications by its use of shared, derived characters, and also by the evolutionary implications of the method. This new Wagner algorithm differed from that of the “simple Wagner algorithm” as follows:

1. Find the pair of OTU’s that differs the most (using equation 1 from above).
2. Compute the advancement index of each taxon from the interval formed between the two initial taxa.
3. Take the taxa with the largest advancement index and add it to the interval via an HTU.
4. Find the next unplaced OTU with the largest “advancement index,” find the interval from which it differs least.

This produces a network, rather than a tree, and does not assume that any of the taxa are ancestral. Farris suggested that the network could be converted into a phylogenetic tree by rooting it at one of the taxa within the tree, or an interval within the network. Completing the process of constructing phylogenies using this method requires that the characters be optimized onto the tree.

The earliest programs implementing the Wagner algorithm did not necessarily find the most parsimonious tree for large data sets. The program needed to run multiple times and have a method of comparison in order to determine whether it has indeed found the shortest tree, or if there were multiple equally parsimonious trees. In a large matrix, examining every possible tree could require an enormous amount of computer time, and thus it became necessary to develop heuristic methods to try to find the shortest tree. Today’s parsimony programs, such as those in PAUP, Hennig 86, and NONA, use a variety of heuristic algorithms to rerun the data to attempt to ensure that the most parsimonious tree or trees are found. For small numbers of taxa and characters, the Branch and Bound algorithm (Hendy and Penny 1982), which guarantees finding the shortest tree, or the Exhaustive Search option, which enumerates all possible trees, can be employed.

As phylogeneticists began to analyze increasingly larger and more complicated data sets, shortcomings in the original computer programs became evident. In the decade following Farris’ (1970) contribution, a number of algorithms were developed, such as Fitch parsimony (Fitch 1971) and Dollo parsimony (Farris 1977), which were incorporated into the existing programs as alternatives to Wagner parsimony. These differed primarily in their assumptions and restrictions regarding character evolution and are discussed in more detail by Wiley et al. (1991).
The first iteration of the Wagner algorithm did not take into account multistate characters, and therefore technically it was not possible to have unordered states, since polarized binary characters are automatically ordered. Initially, before more variations were developed for the algorithm, it was suggested that all multiple character states be divided into multiple binary characters [e.g., a single multistate transformation series of an imaginary character (absent (0), short (1), long (2)) would be divided into two separate characters (absent (0), present (1)) and (short (0), long (1))]. Current algorithms allow for multistate transformation series and allow characters to be run either polarized or unpolarized, and either ordered or unordered, at the discretion of the user. Again, the advantage of phylogenetic methodology is that these decisions are transparent (if they are reported) and repeatable; with the same data set, anyone can rerun an analysis using the same settings to check the reliability of the analysis, or change the settings to see if the results are different.

Whatever algorithm you use to build a tree, in most cases some characters will not be decisive at every node (Farris 1970). It is therefore important for the purpose of studying character evolution to be able to optimize characters on a tree. There are two types of optimization, ACCTRAN (Farris 1970; Wiley et al. 1991) and DELTRAN (Swofford and Maddison 1987; Wiley et al. 1991). The ACCTRAN setting accelerates the transformation of a character on a tree, pushing the evolution toward the root. This is equivalent to preferring parallelisms to reversals, if the choice does not affect the tree length. DELTRAN delays the transformation of a character on a tree, essentially choosing reversals over parallelisms when they are equally parsimonious (Wiley et al. 1991). When there are no equally parsimonious alternatives, both ACCTRAN and DELTRAN will provide the same result (Figures 5.4 and 5.5).

Figure 5.4
DELTRAN tree (length = 9 steps). Redrawn and modified from Wiley et al. (1991)
5.3.2 Development of outgroup comparison

As noted above, Wagner algorithm generates a minimum-length network (sometimes called an “unrooted tree”). In order to convert a Wagner network into a phylogenetic tree, the network must be rooted in some manner. Increasingly, published studies convert the network into a tree by rooting it with an arbitrarily chosen single taxon not included in the group being analyzed (called the ingroup). This protocol should not be mistaken for the method of outgroup comparisons that emerged in phylogenetics during the 1970s. The distinction is slight, but significant, and must be understood in light of Hennig’s perspective on the issue of ancestors.

Hennig objected strongly to the notion that phylogeny reconstruction could be achieved by reconstructing a series of archetypal ancestors, from which particular descendant species could be derived. His position was that each species was a unique mosaic of plesiomorphic and apomorphic traits. Archetypes, defined as ancestral species exhibiting only plesiomorphic traits, thus did not exist; therefore, no single taxon could be used to determine the plesiomorphic and apomorphic traits for any analysis. Or, using current jargon, rooting a network with a single outgroup taxon is sufficiently robust in the Hennigian system only if that taxon is the archetype ancestor of the ingroup, something the Hennigian system disavows.

As can be seen from the discussion above, the early development of the Wagner algorithm was not informed directly by Hennigian reasoning. Rather, it relied on the groundplan-divergence method, based on a priori recognition of an

![Diagram](attachment:image.png)
archetypal ancestor. When Farris (1970) abandoned the a priori reliance on an ancestor, the Wagner algorithm reverted to a method for producing an unrooted network. Lundberg (1972) made a significant contribution to linking the results of Wagner analyses with Hennigian analyses by differentiating ancestors from outgroups.

He developed a method to determine an ancestor from within a network from the data within that same network. He opined that the structure of a network makes certain character states more likely to be ancestral, helping to determine which interval should form the root of the tree of a parsimony-optimized network. The transition of emphasis from searching for ancestors to identifying outgroups was critical in linking Wagner with Hennig.

The idea that similarity in traits even among distantly related species was due to homology (i.e., plesiomorphy) rather than independent evolution (homoplasy) was established before the development of Hennigian systematics.

> ... it would in most cases be extremely rash to attribute to convergence a close and general similarity of structure in the modified descendants of widely distinct forms. The shape of a crystal is determined solely by the molecular forces and it is not surprising that dissimilar substances should sometimes assume the same form; but with organic beings we should bear in mind that the form of each depends on an infinitude of complex relations, namely on the variations that have arisen, these being due to causes far too intricate to be followed out,—on the nature of the variations that have been preserved or selected, and this depends on the surrounding physical conditions, and in a still higher degree on the surrounding organisms with which each being has come into competition,—and lastly, on inheritance (in itself a fluctuating element) from innumerable progenitors, all of which had their forms determined through equally complex relations. It is incredible that the descendants of two organisms, which had originally differed in a marked manner, should ever afterwards converge so closely as to lead to a near approach to identity throughout their whole organisation. If this had occurred, we should meet with the same form, independent of genetic connection, recurring in widely separated geological formations; and the balance of evidence is opposed to any such admission. Darwin (1872 pp 127–128)

Despite the fact that the connection between outgroups and the Auxiliary Principle had been around for a long time, there was no codification until the late 1970s. Engelmann and Wiley (1977) were the first to provide a rationale for outgroup comparisons. They pointed out that the reference to species outside the ingroup permits a researcher to distinguish traits that truly conflict with phylogeny (homoplasies) from those that only appear to conflict (plesiomorphies). This in turn creates the possibility that phylogenetic analysis could become testable, at
least with respect to Darwinian concepts. Watrous and Wheeler (1981) expanded on this idea, suggesting a number of rules to determine ancestral states for each independent character on the basis of comparisons with an outgroup taxon. The first algorithm to determine ingroup relationships with reference to multiple outgroups was presented by Maddison et al. (1984), who showed that the most robust outgroup comparisons relied on two or more paraphyletic outgroups. This algorithm is incorporated in the program PAUP to root networks when outgroups are specified.

Closely related to the issue of using outgroups to reconstruct ancestral character states are the terms and meanings of “global” and “local” parsimony, which were first applied by Maddison et al. (1984). They proposed a two-step procedure that measures parsimony locally among the outgroups to determine ancestral states and given that these ancestral states then measures locally within the ingroup. This results in one or multiple ingroup cladograms that are most parsimonious globally, i.e., most parsimonious in the context of related groups.

We previously addressed the connection between the Auxiliary Principle and epistemological parsimony. Linking the Auxiliary Principle to outgroup comparisons thus provides a connection, through the Auxiliary Principle, between outgroup comparisons and parsimony. It is the use of outgroups to root the shortest network that makes the Wagner algorithm Hennigian, accounting for high degrees of consistency between Wagner algorithm, groundplan-divergence method, and Hennig argumentation of the same data (Churchill et al. 1984).

5.4 Evaluating the robustness of a parsimony analysis

5.4.1 Character evaluation

As noted in the introduction, Hennigian phylogeneticists are preoccupied with assessing the empirical robustness of their results. There are various methods available to accomplish this goal. These “goodness of fit” measures are useful indicators of the degree of internal conflict among the data (characters) used. Measuring the robustness of the characters and knowing how they behave over the tree topology is one useful approach.

The simplest summary statistic is the tree length; it is merely the number of steps required to produce a particular topology, and it is calculated by adding the number of character changes over the tree (Wiley et al. 1991). Parsimony analysis chooses the tree, or trees with the shortest overall length, given a set of characters.

Consistency indices (Kluge and Farris 1969) attempt to quantify the amount of homoplasy on a particular tree. The original form of the consistency index (CI)
is the ratio of the total number of apomorphic states to the tree length. A high CI indicates there is little homoplasy (i.e., the tree length approaches the minimum number of steps required) and a low CI indicates there is a high degree of homoplasy. This measure is independent of a particular data set and thus can be used to compare trees produced by different data. However, the CI can be inflated by autapomorphies, which do not represent tests of relationships and thus are not informative with respect to the robustness of the tree. Farris (1989) therefore proposed the rescaled consistency index (RC). The RC is an adjusted version of the CI with the influence of characters that do not change the fit of the tree (e.g., autapomorphies) removed. It still gives a relative measure of the degree of homoplasy on a particular tree topology.

5.4.2 Tree evaluation

A second type of evaluation assesses the robustness of the tree topology itself. Decay analysis (Bremer 1988) determines the number of steps required to collapse nodes. To perform a decay analysis, we increase the tree length by successive steps. This shows how many trees exist that are one or more steps longer than the most parsimonious tree (MPT); if there are a number of trees of similar length to the MPT, but with different topologies, we might place less confidence in the MPT. A decay analysis will also reveal how many added steps it takes to collapse individual nodes, and which specific characters influence those nodes. This in turn allows us to test whether a set of functionally correlated characters influences a particular node. TreeRot (Sorenson 1999) is a computer program that uses PAUP* to perform decay analysis, although it is possible (albeit time consuming) to do it manually by successively adding one to the tree length in the search parameters.

Hennigian analysis of a data set may produce more than one tree with the same number of fewest possible steps, a phenomenon known as multiple most parsimonious trees. With a large number of taxa and characters, especially if they contain large amounts of homoplasy or missing data, parsimony frequently generates multiple MPTs. In these cases, it is not possible to designate a single preferred tree; however it is possible to generate a variety of consensus trees to delineate similarities in topologies of different MPTs (Adams 1972; Wilkinson 1994).

There are different techniques to build consensus trees that combine the topological information from two or more trees to create a new summary tree. Strict consensus trees only include monophyletic groups that appear in all of the input trees, and thus usually result in a number of polytomies (Sokal and Rohlf 1981).
Adams consensus trees are slightly more inclusive, purporting to give the most resolution between a set of trees (Adams 1972). These may, however, produce groupings not found in any of the input trees. Majority-rule consensus trees are perhaps the most lenient and frequently used summary tree technique (Margush and McMorris 1981). They are created by building a new tree that contains all monophyletic groups that are supported by a majority of the set of input trees. This means that they may be logically inconsistent with the information produced by one of the MPTs. Consensus techniques are useful as visual summaries of points of agreement or logical consistency between MPTs, but they are not phylogenies, and they are not equivalent to what is produced by phylogenetic analysis of a data matrix. Occasionally, the consensus tree will be one of the MPTs in which case it is a summary tree as well as a phylogeny. Usually, the creation of consensus trees results in the creation of a number of polytomies, or nodes in which the relationships between taxa are unresolved. These are known as soft polytomies when they are created by lack of resolution due to insufficient data or methods. Hard polytomies are actual speciation events, in which a population divides simultaneously into three or more descendent species, or in which two sister species hybridize, forming a third species (Brooks and McLennan 2002). These are impossible to distinguish from soft polytomies with a phylogeny alone.

Bootstrap and Jackknife analyses attempt to estimate the degree of sampling error in the original data set, by attempting to place confidence intervals on phylogenies by making inferences about the variability in the data set. Bootstrapping (Felsenstein 1985a) samples the data set with replacement, that is, it allows for some characters to be sampled more than once, and some not to be included at all and constructs a new data set with the same number of characters. PAUP* (Swofford 1998) and other programs construct a series of these, and build a majority-rules consensus tree that summarizes the results of the resampled data. The number of times a particular group is included in the set of trees that form the consensus is an estimate of the reality of that group, in that the process has measured the amount of variation between the newly sampled data sets. The bootstrap, then, is a measure of the confidence we can place in each node of the tree, like the decay index. Felsenstein (1985a) suggested that a bootstrap value of 95% or greater offers statistically significant support for a clade.

There are a number of caveats of which we must be aware before placing too much faith in the numbers generated by this analysis. Since the bootstrap measures the variation in one set of data, it does not allow us to choose between trees built from different data sets. Felsenstein (1985a) stipulated that a bootstrap assumes characters are independent and equally distributed. He was explicit that the bootstrap indicates repeatability of an analysis given the data, and should not
imply the phylogenetic accuracy of a tree (Soltis and Soltis 2003). It will also be affected by biases such as long-branch attraction (Swofford 1998).

The jackknife is another mode of evaluation, similar to the bootstrap in that it estimates variability in the data set. It is a procedure to resample data by deleting a certain number of characters [either half (Felsenstein 1985a) or another fraction (Farris et al. 1996)] and resampling the data without allowing characters to be duplicated. Characters are randomly and independently deleted from the original matrix to create a new “resampled” matrix, and like the bootstrap, many matrices are produced and the results are compiled into a consensus tree. For a review of the different kinds of jackknife resampling (delete-half, parsimony, weighted) and the assumptions and problems with each see Efron (1979), Wu (1986), and Farris et al. (1996).

5.5 Ontological Parsimony: Maximum likelihood and Bayesian likelihood

5.5.1 A precis of maximum likelihood in phylogenetics

Microbiology made significant progress in the late 1950s when the first proteins were sequenced. Molecular data were soon realized to be an important source of phylogenetic information useful in inferring evolutionary relationships (Sneath and Sokal 1973; Neyman 1974). Edwards and Cavalli-Sforza (1963, 1964) first explored the idea that likelihood could be applied to phylogeny reconstruction. Edwards and Cavalli-Sforza (1964; Cavalli-Sforza and Edwards 1967) later described a likelihood method for phylogenetic inference using blood-group allele frequency data in human populations. Neyman (1974) was the first to apply likelihood analysis to nucleotide sequences, and presciently suggested that this approach might become important in the future. Farris (1973) and Felsenstein (1973) published likelihood algorithms for phylogeny reconstruction, however problems of computational difficulties continued to limit a likelihood method for phylogenetic inference to the theoretical rather than practically operational. Felsenstein (1981b) introduced the first computationally efficient maximum likelihood algorithm for discrete character nucleotide sequence data. Just as the Wagner algorithm became the algorithm of choice for quantitative Hennigian analyses, nearly all phylogenetic applications of maximum likelihood are adapted from Felsenstein’s early work. Since then, maximum likelihood methods have become increasingly popular in phylogenetic studies (Swofford et al. 1996; Huelsenbeck and Crandall 1997; Tuffley and Steel 1997; Felsenstein 2004). These approaches are most commonly used in molecular phylogenetics (Swofford
et al. 1996; Huelsenbeck and Crandall 1997; Huelsenbeck et al. 2002; Ronquist 2004), but morphology-based and combined likelihood and Bayesian methods have been proposed and are being refined (Lewis 2001; Nylander et al. 2004; Ronquist 2004).

### 5.5.2 Likelihood methods

Several methods for inferring phylogenies from nucleotide sequence data are available, resulting in an often-heated debate among evolutionary biologists over the “best” way to approach phylogeny reconstruction (Goldman 1990; Penny et al. 1992; Swofford et al. 1996; Huelsenbeck and Crandall 1997; Steel and Penny 2000). Maximum likelihood methods evaluate a hypothesis of evolutionary relationships using a presumed model of the evolutionary process and evaluate the probability that it would give rise to the observed data, which is typically DNA sequences of the terminal taxa (Felsenstein 1973, 1981b, 2004; Swofford et al. 1996; Huelsenbeck and Crandall 1997). It should be noted that there are several different types of the likelihood (Steel and Penny 2000; Goloboff 2003). Phylogenetic-likelihood approaches use maximum average likelihood, a form of maximum relative likelihood (except Farris 1973, that adopted evolutionary pathway likelihood), and only this form applies to the discussion below (Steel and Penny 2000).

The likelihood of a hypothesis (Fisher 1922) is the probability, \( P \), of the data \( D \), given the hypothesis \( H \):

\[
L = P(D|H) \tag{3}
\]

The likelihood of a parameter is proportional to the probability of the data and it gives a function that usually, but not always, has a single maximum value, which Fisher called the maximum likelihood. The likelihood does not estimate the probability of the hypothesis, which is assumed to be true in the likelihood formulation (Eq. 3).

Likelihoods are calculated for possible tree topology, given the data and assuming a particular model of molecular evolution (Felsenstein 1973, 1981b, 2004; Swofford et al. 1996). In the likelihood equation, Eq. 3, the hypothesis, \( H \), contains three distinct parts: (1) a mechanism or model of sequence evolution, (2) a tree or a hypothesis of relationships, and (3) branch lengths (Penny et al. 1992). For a given data set, likelihoods are calculated for each of the possible tree topologies, or a sample of them, and the tree topology with the highest overall likelihood is the preferred phylogenetic hypothesis. The number of possible tree topologies increases with the number of terminal taxa included in the analysis.
This can be computationally laborious if the data set is large, and especially if the maximum likelihood model uses rooted trees in its calculus. However, the most general and most commonly used models in molecular analyses are time reversible (Rodriguez et al. 1990; Swofford et al. 1996). With a time reversible model the probability of character state change from state \( i \) to state \( j \) is the same as the probability of state change from state \( j \) to state \( I \) (Felsenstein 1981b). Under this condition the likelihood of the tree does not depend on the position of the root, and the use of unrooted networks greatly reduce the total number of trees to be evaluated, and decrease computation time (Rodriguez et al. 1990; Swofford et al. 1996). In a four-taxon statement, there are three possible unrooted networks (Figure 5.6):

![Figure 5.6](image)

**Figure 5.6**
The three potential unrooted networks of the four-taxon, four-character statement

<table>
<thead>
<tr>
<th>Site</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taxon</td>
<td>A</td>
<td>C</td>
<td>A</td>
<td>C</td>
</tr>
<tr>
<td>Taxon B</td>
<td>G</td>
<td>G</td>
<td>A</td>
<td>C</td>
</tr>
<tr>
<td>Taxon C</td>
<td>C</td>
<td>G</td>
<td>G</td>
<td>C</td>
</tr>
<tr>
<td>Taxon D</td>
<td>C</td>
<td>A</td>
<td>C</td>
<td>G</td>
</tr>
</tbody>
</table>

For each network, likelihoods are evaluated on a site-by-site basis. The probabilities (\( P_{ef} \)) of all possible combinations of character states at the internal nodes are calculated for each site (i.e., each character) using a specified model of molecular evolution and estimated branch lengths (see below). The likelihood of the site is the sum total of all the probabilities for all the possible combinations of character states at the internal nodes of the network (Felsenstein 1981b; Swofford et al. 1996; Huelsenbeck and Crandall 1997). In the four-taxon case, there are two internal nodes and therefore 16 possible combinations of character states and 16...
probabilities for each site (when gaps are not considered) (Eq. 4). It is clear that some of these combinations, or scenarios, are more plausible than others given the data at the leaves. However, every combination of states at the internal nodes is theoretically possible and each is therefore given a probability (Figure 5.7).

\[ L_{\text{site}} = P_{AA} + P_{AC} + P_{AG} + P_{AA} + P_{CA} + \cdots \]  

(4)

Figure 5.7
Illustration of the possible character states at the internal nodes for the four-taxon, four-character statement

<table>
<thead>
<tr>
<th>Taxon</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>A</td>
<td>A</td>
<td>C</td>
<td>A</td>
</tr>
<tr>
<td>B</td>
<td>G</td>
<td>G</td>
<td>A</td>
<td>C</td>
</tr>
<tr>
<td>C</td>
<td>C</td>
<td>G</td>
<td>G</td>
<td>C</td>
</tr>
<tr>
<td>D</td>
<td>C</td>
<td>A</td>
<td>C</td>
<td>G</td>
</tr>
</tbody>
</table>

The overall likelihood of the network or tree is calculated as the product of the site likelihoods \(L_{\text{site}}\) because each site is assumed to evolve independently of one another (Eq. 5).

\[ L_{\text{tr1}} = L_{\text{site1}} \times L_{\text{site2}} \times L_{\text{site3}} \times \cdots \times L_{\text{sitenn}} \]  

(5)

Theoretically, this procedure is repeated for each possible unrooted network, although there are “pruning” algorithms that do not require every tree to be considered thereby improving computational efficiency in large data sets (Felsenstein 1981b; Swofford et al. 1996). The network with the highest overall likelihood is the maximum likelihood result, and the preferred phylogenetic hypothesis. It is the network topology that maximizes the likelihood function for the data given the specified model (Felsenstein 1973). It is possible that the network represents only a local maximum, or that it is one of a larger number of equally likely networks. It is important to remember that the likelihood of a network is
NOT \( L = P(H|D) \), which is the probability of the hypothesis. The network is converted into a tree by rooting it with an outgroup or a molecular clock (Swofford et al. 1996; Felsenstein 2004).

### 5.5.3 Models of molecular evolution

Likelihood analyses involve similar assumptions about the evolutionary process as other methods, including that evolution occurs in a branching pattern and is independent in different lineages (Swofford et al. 1996). The probability of a particular combination of character states at the internal nodes of the unrooted network (e.g., Eq. 4.), are calculated using a specified model of molecular evolution, which requires further assumptions about the nucleotide substitution process, including that sequence evolution can be modeled as a random, or stochastic, process (Rodriguez et al. 1990). Substitution models are typically based on a homogeneous Markov process (Rodriguez et al. 1990; Swofford et al. 1996) that assume the probability of a state change at one site does not depend on the history of that site and that probabilities of substitution do not change significantly in different parts of the tree (Felsenstein 1981b, 2004; Swofford et al. 1996).

A DNA substitution model is expressed as a table of rates (substitutions per site per evolutionary distance unit) at which nucleotides are replaced by alternate nucleotides known as the Q matrix (Rodriguez et al. 1990; Swofford et al. 1996; Huelsenbeck and Crandall 1997). The term \( Q_{ij} \) in the instantaneous rate matrix, \( Q \), represents the rate of change from base \( i \) to base \( j \) over an infinitesimal evolutionary time period \( dt \) (Swofford et al. 1996). The Q matrix in Eq. 7 represents the general time reversible (GTR) model. Change probabilities are determined by the relative rate of change supported by the instantaneous rate matrix \( Q \) and estimated branch lengths.

\[
Q = \begin{bmatrix}
-\mu(a_{fC} + b_{fG} + c_{fT}) & \mu a_{fC} & \mu b_{fC} & \mu c_{fT} \\
\mu a_{fC} & -\mu(a_{fA} + b_{fG} + c_{fT}) & \mu b_{fG} & \mu c_{fT} \\
\mu b_{fA} & \mu a_{fC} & -\mu(b_{fA} + d_{fC} + f_{fT}) & \mu f_{fT} \\
\mu c_{fA} & \mu e_{fC} & \mu f_{fC} & -\mu(c_{fA} + e_{fC} + f_{fG})
\end{bmatrix}
\]

(6)

The rows and columns correspond to the bases A, C, G, and T. The mean instantaneous rate factor, \( \mu \), is modified by the relative rate parameters \( a, b, c, \ldots, f \). The relative rate parameters correspond to each possible base substitution scenario. The rate parameter is the product of \( \mu \) and the relative rate parameters. The remaining terms (\( f_A, f_C, f_G, f_T \)) represent the frequencies of the four bases.
The base frequency parameters, calculated over all taxa, are assumed to be at equilibrium and that the rate of change from one base to another is proportional to the equilibrium frequency (Rodriguez et al. 1990; Swofford et al. 1996).

The rates defined in the Q matrix are per instant of time $dt$, in order to calculate the likelihoods of each site, the probabilities $(P_{ij})$ of the possible state changes along a branch length of $t$ (Swofford et al. 1996) must be determined. A substitution probability matrix (sensu Swofford et al. 1996) is calculated using Eq. 7:

$$P(t) = e^{Qt}$$

Calculation of the exponential is complex, involving the decomposition of the Q matrix into is eigenvectors and eigenvalues (see Swofford et al. 1996 for references). For the simple Jukes–Cantor model, these values are relatively easily evaluated because there are only two probabilities, the probability of a state change and the probability of stasis, such that the transition probability matrix consists of two values, as in Eq. 8:

$$P_{ij}(t) = \begin{cases} 0.25 + 0.75 e^{-\mu t} (i = j) \\ 0.25 - 0.25 e^{-\mu t} (i \neq j) \end{cases}$$

The substitution probability matrix that corresponds to the GRT instantaneous rate matrix Q in Eq. 7 similarly has 12 values, 1 for each of the different substitution rates. The branch lengths are unknown prior to the analysis and must be estimated in the course of the likelihood calculation (Goloboff 2003). Estimation of branch lengths involves an iterative algorithm in which each branch is optimized separately (Felsenstein 1981b; Swofford et al. 1996). Unlike the rate and frequency parameters, branch lengths are specific to a particular tree topology. For each tree, multiple different branch lengths need to be evaluated, and branch lengths must be recalculated for each unrooted tree considered (Penny et al. 1992).

Models employed in likelihood analyses make explicit assumptions regarding sequence evolution, the number of which depends on the particular model of sequence evolution used in the particular analysis (Swofford et al. 1996). The general time reversible model is the most general stochastic model of nucleotide substitution. It models base substitution as a random Markov process in which substitution rates are independent among sites, constant in time, equal in two lineages, and that the ancestral sequence base frequencies represent the equilibrium frequencies (Rodriguez et al. 1990). The GTR model has a maximum of 12 different substitution rates (estimated from the data and using the
aforementioned assumptions in their calculus) as expressed in the above Q matrix (Eq. 7) and at least seven parameters (Rodriguez et al. 1990). Because of its greater complexity, nearly all models (including JC, K2P, K3ST, L, TK, GIN, and TN) can be considered a special case of the GTR model when it is restricted accordingly (Rodriguez et al. 1990). For example, the Jukes–Cantor (often abbreviated JC69) model is the simplest model and assumes that all base substitutions are equally likely (i.e., all rate parameters are equal) and that the base frequency parameters are equal \( f_A = f_C = f_G = f_T \) in the above Q matrix). The K2P model has two rate parameters since it considers differences in rates between transition and transversion type substitutions (Rodriguez et al. 1990). The K3ST model considers three substitution rates, one for transitions and two for each of two types of transversions.

The mathematical procedures of likelihood have several dependencies that are not met in the calculus of maximum likelihood methods. The summation of the probabilities in Eq. 5 does conform to the disjunction criterion of the calculus (Siddall and Kluge 1997). For probabilities to be multiplicative, the change probabilities must be independent. The base compositional frequency parameters in the Q matrix (Eq. 7) are derived from the terminal taxon base sequences over all characters in the analysis (Siddall and Kluge 1997). Rate parameters (relative rate and mean rate) are calculated using the Q matrix (Eq. 6) and the assumption that base frequencies remain constant over evolutionary time (Rodriguez et al. 1990). All sites use the Q matrix to calculate the \( P_{ij} \) values and therefore the probability of any character is dependent on all other characters through the frequency parameters in Q. Characters and their associated probabilities are not independent quantities but are assumed to be in the calculus of the method (Felsenstein 1973, 1981b; Rodriguez et al. 1990; Swofford et al. 1996). The nonindependence of site change probabilities may be one factor responsible for the fact that the total likelihood of the universe of possible trees does not sum to unity (Felsenstein 1981b). The true probabilities for character changes should be calculated on an individual basis because they are connected with unique and historically contingent events (see below), as correctly pointed out by Farris (1973). But this is clearly impossible, as it not only requires knowledge of the true history before undertaking an analysis, it also requires an objective and consistent way of determining the probability of a novel, context-specific evolutionary event (see below, Farris 1973). It would also be computationally impossible (Felsenstein 1973, 1981b; Siddall and Kluge 1997). As currently and commonly employed in phylogenetic maximum likelihood methods, frequency probability theory is technically violated (Yang 1996; Siddall and Kluge 1997).
5.5.4 Choosing a model: more ontological parsimony

The choice of an appropriate model is a critical aspect of a phylogenetic-likelihood analysis. Posada and Crandall (1998) argued that the use of their ModelTest program makes selection of the model objective. There are many models of molecular evolution, and the determination of which to use can significantly influence the results of an analysis. Models range in complexity from the relatively simple Jukes–Cantor model through the most complex GTR model. Currently there are at least 16 models that are commonly used in molecular systematics, most of which are simply special cases of the GTR model (Rodriguez et al. 1990). Each of the 16 basic models is varied with $G$ (gamma distribution), $I$ (proportion of invariable sites), and both ($G+I$) for a total of 56 different models (Posada and Crandall 1998). ModelTest ranks them in terms of their relative complexity (i.e., more complex = more parameters). The overall likelihood score of a tree increases with increasing complexity of the model, but the accuracy of the model decreases due to the increased number of estimated parameters (Huelsenbeck and Rannala 1997a). The program conducts pairwise comparison of the maximum likelihood trees generated under each model using hierarchical-likelihood ratio tests (Huelsenbeck and Crandall 1997; Huelsenbeck and Rannala 1997a; Posada and Crandall 1998; Johnson and Omland 2004). When no statistically significant difference between two trees is found, the simplest model of these is selected. Model selection based on the relative likelihood values is an ontological appeal to the principle of parsimony, because choosing the least complex explanation of the data rules out the possibility that evolution proceeded in a more complex manner (Huelsenbeck and Rannala 1997a).

5.5.5 Criticisms of the models

Many criticisms of maximum likelihood methods are directed at this a priori dependence on a model. Evolutionary realism of the models employed in likelihood analyses is often compromised by approximations that are designed to improve the computational efficiency of the algorithms. For example, Lockhart et al. (1994) suggested that a modified GTR model, in which time-reversibility is relaxed, across site rate variation is considered, and the nucleotide compositional frequencies are flexible, allows more evolutionary “freedom” than any other model, and best considers the historical ambiguity and evolutionary contingency of the evolutionary process. They suggested that this complex, parameter rich, and computationally intensive model should be logically preferred over all other
models, if inferring phylogeny using the most realistic conception of evolution (i.e., evolution is complex) is the goal of the analysis. The problem is that relaxing the time-reversibility assumption introduces the need for rooted trees and is accompanied by additional computational problems (Swofford et al. 1996). Relaxing the assumption that rates are equal across all sites can be accomplished by adding another relative rate parameter to the matrix, which commonly involves modeling rate heterogeneity using the Gamma distribution (Swofford et al. 1996). If this distribution is modeled as continuous (as it should be), it again becomes computationally laborious, and a discrete distribution typically serves as a computationally more efficient approximation (Swofford et al. 1996).

The likelihood method also requires that numerous parameters be approximated using the data, and relies heavily on the frequency parameters that are taken directly from the observed sequences (Swofford et al. 1996). In this sense, likelihood methods require that the processes that maintain systems today were persistent throughout the entire evolutionary history of the clade being investigated (Brooks and McLennan 2002). Siddall and Kluge (1997) and Lockhart et al. (1994) provide an empirical example where the nucleotide frequencies are very different across the terminal taxa, proving that the assumption of equilibrium base frequencies is not at all tenable, and that the assumption of such can affect likelihood phylogenies. When the model is insufficient or inappropriate, appeals to statistical consistency are rendered moot (Siddall and Kluge 1997).

5.5.6 Criticisms of the method

Phylogenetic applications of likelihood rely on two fundamental assumptions: that evolution is independent in different lineages as well as independent in different sites for a given tree, which are essential for the probability calculations on which the method is based (Felsenstein 1981b, 2004; Rodriguez et al. 1990; Swofford et al. 1996). Both assumptions are methodologically problematic because they are unrealistic and/or violated in the calculus (Siddall and Kluge 1997; Huelsenbeck and Nielsen 1999), but likelihoodists appeal to simulations to argue that the method is generally robust to violations of these assumptions (Yang 1994; Swofford et al. 1996). It is also assumed that the same stochastic process of substitution applies in all lineages (Felsenstein 1981b).

and Kluge 1997; Kluge 2001; Goloboff 2003) cite operational and philosophical problems of likelihood methods and discourage its use in phylogenetic inference.

Likelihood methods rely on a specified model of sequence evolution to infer phylogenetic relationships. This is an inductive approach as the assumptions of the model are clearly deterministic to the result of the analysis. In phylogenetic-likelihood analyses, like all inductive approaches in science, all interpretations of the results come with the caveat “if the model is true, then . . . .” Because likelihood is inductive estimation of phylogeny, particular caution must be taken in interpreting the results and to avoid circularity. We may know which of the models best fits the data according to ModelTest, but how can the validity of the model itself be independently tested? Testing of the validity of models, although it has been recognized as important (Goldman 1990), is rarely done in practice (Siddall and Kluge 1997). In a hypothetico-deductive framework, the assumptions of the method are background and not deterministic to the result. Only the observable data is considered, maximizing the explanatory power of the hypothesis and minimizing ad hoc hypotheses (i.e., hypotheses that confide in “nonobservables,” such as long branches in a likelihood framework). It has been argued that many of the simple assumptions of the evolutionary models (i.e., the frequency of transitions versus transversions) constitute grounded knowledge about the process of molecular evolution and therefore is an acceptable background assumption. However, the maximum likelihood approach suggests that this is and has always been the case throughout all situations in the evolution of a group, which is a difficult assertion with respect to the historically contingent nature of the evolutionary process (Siddall and Kluge 1997).

Likelihood methods are based on frequency probability theory. Frequency probability is concerned with prediction of future events (Fisher 1922). The aim of phylogenetic systematics is to discover the unique evolutionary history of a group of organisms, or to elucidate its past (Kluge 1997). A species must be considered a historical entity (Kluge 1990), evolutionary transformations are unique and spatiotemporally restricted historical events (Siddall and Kluge 1997). Frequency-probability based methods of phylogenetic inference, such as maximum likelihood, apply frequency probability to a historical singularity, which is outside of the realm of future-predictive probability theory (Siddall and Kluge 1997). As noted above, likelihood methods assign all trees a nonzero probability, but in reality one tree has a probability of 1.0, and others have a probability of zero. One must be cognizant that maximum likelihood inference of phylogenies is philosophically unsound because it employs frequency probability theory to estimate a nonprobabilistic phenomenon.

In spite of these numerous criticisms, likelihood approaches have become increasingly popular, particularly with molecular phylogeneticists. Recently, other
methods incorporating likelihood logic have become popular in systematics, the most popular of which is Bayesian analysis (Huelsenbeck et al. 2002).

**Bayesian likelihood** Bayesian methods calculate the posterior probabilities of phylogenetic hypotheses (trees) using a version of Bayes’ theorem in which the likelihood of the tree and the prior probability of the tree are considered (Huelsenbeck et al. 2002). Huelsenbeck et al. (2001, 2002) provide excellent recent review of the history and mechanics of Bayesian inference methods in phylogeny.

Reverend Thomas Bayes, living in the early half of the eighteenth century, was an English mathematician who was interested in the concept of using *a priori* knowledge to predict future events. His paper, “An Essay Towards Solving a Problem in the Doctrine of Chances,” published 2 years after Bayes’ death in 1761, introduced what would become known as Bayes’ theorem (Eq. 10; Barnard and Bayes 1958).

\[
P(H|D) = \frac{P(D|H) \times P(H)}{P(D)}
\]

The posterior probability, \([P(H|D)]\), is the probability of the hypothesis given the observations, or data (D). Note that this is different from likelihood, which is the probability of the data given the hypothesis. However, the likelihood, \(P(D|H)\), is a parameter in the calculation of the posterior probability. \(P(H)\) is the prior probability of the hypothesis before the observation, data, or analysis, and reflects the original beliefs regarding the problem. The probability of the hypothesis is updated to take into account the observations, and Bayes’ theorem describes the relationship between the prior and posterior probabilities. It was not until the latter half of the twentieth century that Bayes’ ideas would be applied to the inference of phylogenies. Felsenstein (1968) briefly discussed Bayesian ideas as they could apply to phylogeny reconstruction in his Ph.D. thesis, but the statistical and computational framework with which to derive reliable approximations of posterior probabilities was not available at the time (Huelsenbeck et al. 2002).

Three independent groups introduced Bayesian methods similar to those currently in use in 1996 (Li 1996; Mau 1996; Rannala and Yang 1996). Bayesian methods to estimate ancestral character states have been developed (Pagel et al. 2004; Ronquist, 2004).

Huelsenbeck et al. (2001, 2002) provide overviews of the Bayesian methodology. Bayesian phylogenetic inference evaluates phylogenetic hypotheses with the posterior probabilities of trees. The posterior probability of each tree is calculated using the following Bayes-based equation (Eq. 11), where the tree topology (including branch lengths) is the hypothesis, and the data is typically molecular sequences of the terminal taxa in the analysis.
The likelihood parameter, $P(\text{Data}|\text{Tree})$, is calculated using the same general methodology and same models of molecular evolution described above for the maximum likelihood approach. However, one major difference between Bayesian and maximum likelihood methods is that Bayesian likelihood calculation not only involves summation over all possible combinations of model parameters and branch lengths but also includes a prior probability density distribution of these latter variables (Huelsenbeck et al. 2002). The prior probability of the tree, $P(\text{Tree})$, is usually considered to be equal for all trees a priori (Huelsenbeck et al. 2001). The use of equal priors implies that no particular topology is preferred over any other a priori and eliminates the sometimes difficult task of calculation of complex priors when hypotheses vary with respect to their preconceived probabilities. However, the prior for any given tree or set of trees can be set to reflect researcher experience, the results of previous analyses, or taxonomy (Huelsenbeck et al. 2002). The denominator, simplified here as $P(\text{Data})$, is a normalizing factor that involves summation over all trees (Huelsenbeck et al. 2002). The posterior probability, $P(\text{Tree}|\text{Data})$, can be viewed simply as the probability that the tree is “correct,” given the data, the priors, and that the model of character change used in the likelihood calculation is correct (Huelsenbeck et al. 2002). There are several ways to present the results of a Bayesian analysis. The tree with the maximum posterior probability can be selected as the preferred phylogenetic hypothesis, this is also known as the MAP, maximum a posteriori, estimation of phylogeny (Huelsenbeck et al. 2002). Another method is to construct a 95% credibility consensus tree by starting with the MAP and consecutively adding the next most probable trees until the probabilities total 0.95 (Huelsenbeck et al. 2001, 2002). The method preferred by Huelsenbeck et al. (2002) is to present a summary of the results on the MAP or another consensus tree, as is typically done with bootstrap.

5.5.7 Posterior probability estimation using Markov chain Monte Carlo

Calculation of the posterior probability of a tree is computationally expensive because it involves summation over all possible trees, and for each tree requires integration over all possible permutations of branch lengths and substitution-model parameters (Huelsenbeck et al. 2001, 2002). This is not possible in most practical applications due to computational and time constraints, and requires
that posterior probabilities be approximated (Huelsenbeck et al. 2002). Markov chain Monte Carlo (MCMC) methods are used to approximate the posterior probabilities of trees, and allow contemporary Bayesian methods to be computationally feasible (Tierney 1994; Huelsenbeck et al. 2001, 2002). The MCMC method is used to sample the posterior probability distribution of trees. The application of the MCMC to phylogeny inference is discussed in detail by Mau and Newton (1997), Yang and Rannala (1997), Mau et al. (1999), Larget and Simon (1999), Newton et al. (1999) and summarized in Huelsenbeck et al. (2001, 2002) and Pagel et al. (2004). The general process MCMC is used to approximate the posterior probability density is as follows. First, a random tree is selected. Second, another tree is proposed by changing one variable of the original tree (e.g., topology, branch length, model parameters, etc.), and the two trees are compared using the Metropolis–Hastings algorithm (Metropolis et al. 1953; Hastings 1970; Green 1995; Huelsenbeck et al. 2002). If the tree represents an improvement, it is accepted, or sampled. If not, the tree is either accepted or rejected proportional to the likelihood ratio between it and the previous tree (Pagel et al. 2004). The chain stabilizes after a sufficient period of run time (called the “burn-in”). Once stable the chain randomly walks through the universe of trees, sampling each tree in proportion to its frequency in the actual posterior density (Pagel et al. 2004). The longer the chain is run, the greater precision with which the actual posterior distribution of trees is approximated (Pagel et al. 2004). Metropolis-coupled MCMC uses multiple, simultaneous Markov chains, improves mixing and convergence, and allows exceedingly large data sets that are beyond the scope of conventional single-chain MCMC Bayesian methods to be analyzed (Geyer 1991; Huelsenbeck et al. 2001, 2002).

5.5.8 Advantages of Bayesian likelihood

A major advantage of the Bayesian method is the ease with which posterior probabilities can be interpreted (Huelsenbeck et al. 2002). Under the assumption that the evolutionary model is true and that the MCMC has accurately sampled the posterior probability distribution, the posterior probability value represents the probability that the tree is correct given the data and the priors. Similarly, the proportion of trees in the MCMC sample in which a monophyletic group appears represents the probability that the clade is “true,” given the caveats of priors, model, and data.

As in maximum likelihood analyses, the result of the Bayesian analysis is dependent on the model of sequence evolution being “correct.” Bayesian approaches to phylogeny require a likelihood value of a given tree topology for
their calculation of the posterior probability of that evolutionary scenario. The likelihood parameter in the Bayesian method uses the same models and their associated assumptions as the maximum likelihood methods described above, and the caveats inherent in maximum likelihood phylogeny estimation with respect to evolutionary models also apply to Bayesian analysis (see above discussion of likelihood criticisms).

Computational and time constraints require that the posterior probability distribution be approximated using MCMC techniques (Huelsenbeck et al. 2001, 2002). Chains may fail to provide an accurate estimate of posterior probability distributions if they are not allowed to run long enough, or if mixing is a problem due to widely separated peaks in the distribution. It is difficult to know when a chain has run long enough to provide an acceptable estimate of posterior probabilities. The longer the chain is run, the more precise the estimate of posterior probability distribution. Huelsenbeck et al. (2002) propose three recommendations to ensure that the posterior probability is sampled reliably: (1) run several long chains, and check for consistency in results; (2) run multiple chains, each starting from a random tree and check for consistency; and (3) monitor the model parameters for convergence. The Metropolis-coupling technique promotes good mixing and increases the speed of convergence (Huelsenbeck et al. 2001, 2002).

Some view it as an advantage that Bayesian analysis requires the incorporation of previous knowledge or beliefs in terms of prior probabilities. The mechanics of formulating priors can be difficult if one chooses to base these off the results of previous analyses or taxonomy ("complex priors," Huelsenbeck et al. 2002). Making the prior probabilities of each tree equal eliminates the use of complex priors, as well as any a priori assumptions that any hypothesis is more probable than any other in light of prior beliefs, clearly this approach is not in the true Bayesian spirit.

### 5.5.9 Criticisms of Bayesian likelihood

Bayesian approaches to phylogeny require a likelihood value of a given tree topology (i.e., phylogenetic hypothesis) for their calculation of the posterior probability of that evolutionary scenario. The likelihood used in the Bayesian method requires the same models and their associated assumptions as the maximum likelihood methods described above, and thus all of the cautions inherent in maximum likelihood phylogeny estimation also apply to Bayesian analysis (see above discussion of likelihood criticisms). Analogous to likelihood, the posterior probability of a tree involves summation over all possible trees,
including all their possible permutations in terms of branch lengths and substitution-model parameters (Huelsenbeck et al. 2002). This is impossible to perform in most practical applications due to computational and time constraints, and requires that posterior probabilities themselves must by necessity be approximated. Additionally, Bayesian methods require further assumptions about the prior probabilities associated with each tree. Making the prior probabilities of each tree equal minimizes any a priori assumptions that any topology is more probable than any other.

Harper (1979) had considerable foresight in devising a Bayesian framework aimed toward constructing and testing phylogenetic hypotheses, although this method was largely conceptual and significantly differs from the current practice of Bayesian inference in phylogeny. Likewise, Brooks et al. (1986) proposed the D measure, based on the statistical entropy calculations in information theory. Tribus and McIrvine (1971) suggested that Van der Waals proposed the first link of Bayesian reasoning to statistical concepts of entropy in 1911, when he suggested that the highest entropy state for a system was its a priori most probable state. Lewis (1930) subsequently suggested that increases in entropy were associated with losses of information, and Shannon (1948) used the statistical formulation of entropy as a synonym for expected uncertainty as a foundation of information theory. Kullbach (1951) reinforced the link between information theory and Bayesian reasoning by using the term surprisal (first introduced by R. Levine) as an indicator of departure from the most probable/most expected state. Brillouin (1951, 1953, 1962) showed that there was consistency between Shannon’s use of entropy in information theory and the use of entropy in statistical mechanics and probability theory. Jaynes (1957a, b) derived the formal relationship underlying this consistency and proposed the first entropy maximization principle, in which the maximum entropy state of a system could be formally construed as the most probable state. Parenthetically, and with respect to our discussion of maximum likelihood analyses above, Jaynes’s use of the maximum entropy principle provided a rationale for choosing the most complex, rather than simplest, model for explaining a complex system. He reasoned that adopting the most complex model among all those that explained a system completely would expose our ignorance of possibilities while adopting the simplest would give us a false sense of security, leading us to think we had more complete knowledge than we had.

Gatlin (1975) added to this conceptual framework by discussing two forms of redundancy in the context of biological (specifically DNA sequence) evolution. R-redundancy results from the repeated occurrence of the same symbol in order to get a message across. This is one way to ensure proper communication of a message, but since each symbol must be repeated, R-redundancy is also associated
with reduced message variety. D-redundancy, or Shannon redundancy, results when a single symbol appearing once always subsequently stands for the same thing. D-redundancy is associated with increased message variety, since no symbol need be repeated and thus a larger number of possible messages could be transmitted using the same number of symbols than for situations in which some symbols were repeated. This state may also be associated with reduced message fidelity, however, since missing the symbol initially results in a loss of information subsequently, without hope of a recurrence of the symbol. Gatlin associated D-redundancy with optimal coding in communication systems. Finally, R-redundancy is associated with low information density per symbol (each symbol represents only itself) and D-redundancy with high information density per symbol (each symbol represents many observations).

Brooks (1981) first showed that phylogenetic systematics operationally produced the lowest possible informational entropy configuration for a set of observations over a given set of taxa. This corresponds to the state of greatest difference between $H_{\text{max}}$ and $H_{\text{obs}}$, or the configuration closest to $H_{\text{min}}$ possible given the data. Brooks et al. (1986) later proposed an informal link between this and Gatlin’s D-redundancy, proposing the so-called D-measure for choosing optimal phylogenetic trees on the basis of maximum information density.

Neither Brooks (1981) nor Brooks et al. (1986) related these discussions explicitly to Bayesian reasoning, but there is a clear connection. Following Jaynes (1957a, b), Bayesian approaches in information theory are thus those for which the a priori subjective hypothesis is determined by the entropy maximum principle—the a priori most probable result is the $H_{\text{max}}$, in direct analogy is with the maximum entropy state being the most probable for a closed system. This analogy works in a Bayesian framework if we stipulate that the set of observations we are using in any analysis is a closed subset of all possibilities, i.e., we stipulate that our estimate of $H_{\text{max}}$ is based on a subjective subsample of an imperfectly known universe of characters, and we will not introduce additional observations during the testing procedure.

The entropy maximum is not only analogous to the a priori expected most probable state, it is also the state of lowest information density of the observations, hence least informative, hence least surprising (in a Bayesian sense). For any set of observations (subjective in the sense that they are a subjectively selected subset of all observations, drawn from a universe for which we do not have any sense of the actual size or distribution of variables—even if the observations are obtained objectively), we can objectively compute the most probable state ($H_{\text{max}}$). We can also objectively compute the least probable state ($H_{\text{min}}$), which is the state of greatest information density for the observations, and thus the state of greatest surprise.

198
The most powerful analysis of such data is one that seeks to find the most improbable/highest information density configuration permitted by all the data at hand. For phylogenetic analysis, $H_{\text{max}}$ and $H_{\text{min}}$ can be calculated from the basic data matrix (hence $H_{\text{max}}$ is \textit{a priori}), whereas $H_{\text{obs}}$ is calculated over a set of trees (hence, it is \textit{a posteriori}). The preferred result is the one in which $H_{\text{obs}}$ approaches $H_{\text{min}}$ as closely as possible.

Applying the D measure leads to a number of conclusions for phylogenetics analysis (Brooks et al. 1986): (1) information density is proportional to evolutionary conservatism; (2) dichotomous solutions are preferred over polytomies, as each instance of conservative evolution increases information density; (3) branching topology is irrelevant, there is no \textit{a priori} difference between symmetrical and asymmetrical tree structure in terms of information density, since it is the information that produces the tree, not the reverse; (4) for any data set, the most information dense set of relationships of all taxa over all characters allowed by the data is the shortest tree; and (5) when there are multiple most parsimonious trees, ACCTRAN provides a more information dense summary of the data than DELTRAN.

This version of Bayesian thought applied to phylogenetics, therefore, devolves to straightforward support for Hennigian methods. Missing from this formalism are statistical significance tests. There need to be two of them: first, is the result significantly different from the \textit{a priori} expectation ($H_{\text{max}}$); and second, are less information-dense alternatives (e.g., other equally parsimonious trees or less than most parsimonious trees) for the same set of data significantly different from each other?

### 5.6 How do you decide what method(s) to use?

We have discussed a variety of objective methods for pursuing quantitative phylogenetic analysis. We believe, however, that there are no objective means by which one can choose among these methods. Consequently, it is no surprise that some of the most contentious interactions among phylogeneticists concern the very subjective issue of which methods are “best” or “proper” or “correct.” It is common for groups of contending scientists, faced with such a situation, to resort to philosophical arguments in an attempt to claim priority for one viewpoint over another on the basis of some set of first principles. This has certainly been the case with phylogenetics.

#### 5.6.1 Popper

The first salvo fired in this conflict was by Wiley (1975), who, defending phylogenetic systematics against claims that it was not falsifiable, proposed that
phylogenetic hypotheses of homology could be seen as an exercise in hypothetico-deductive reasoning

> Once a hypothesis of homology is formulated from the world of experience it is tested in two phases: by its own set of potential falsifiers and by a set of potential falsifiers of the phylogenetic hypothesis to which it belongs as a proper subset (i.e., it is tested by other hypotheses of synapomorphy through the testing of the phylogenetic hypotheses which they corroborate). Both phases of testing must be done under the rules of parsimony, not because nature is parsimonious, but because only parsimonious hypotheses can be defended by the investigator without resorting to authoritarianism or apriorism. Wiley (1975) p 236

Hypotheses of homology, together with their connected hypothesis of phylogenetic relationships, can be tested by other independently proposed homologies, which then represent the so-called potential falsifying hypotheses (Popper 1968 p 87). Wiley emphasized that such a process is not circular, but represents a case of “reciprocal illumination,” something Hennig (1966) also claimed. Wiley noted that the preferred phylogenetic hypothesis is the one that has been refuted the least number of times. That Hennigian phylogenetics was justified by the hypothetico-deductive approach of Popper quickly gained support (e.g., especially Gaffney 1979) and still has strong adherents (Kluge 2003) who consider Hennigian phylogenetics to be strongly deductive in nature.

Recently de Queiroz and Poe (2001, see also Faith and Trueman 2001) attempted to link Popperian thought with likelihood approaches, suggesting that likelihood is the basis for Popper’s degree of corroboration. However, for Popper, corroboration was embedded in a falsificationist context, whereas these authors sought to decouple Popper’s ideas about corroboration from those about falsificationism. In this context their degree of corroboration is identical with Popper’s formalization of degree of confirmation (Popper 1997 p 222). This seems to get us nowhere, since it leads back to the position that if a model is accepted as true, or highly typical, its use is justified. But, no objective means is provided for verifying or falsifying the validity of the model beyond the arguments about statistical consistency whose shortcomings we discuss next. Regardless of semantic arguments about corroboration and confirmation, and possibly a high degree of revisionist interpretation of Popper’s views on the relationship between corroboration and falsification, these arguments do not counter the basic observation that maximum likelihood methods are more inductive than deductive in nature. And, if the difference between what we have characterized as the epistemological and the ontological parsimony approaches is the difference between a preference for deduction and a preference for induction, the history of science tells us that there is no objective means for choosing between them.
However popular it has been among some systematists, this battle of philosophical perspectives has been subject to extensive criticism by philosophers, best summed up by Sober (1988), who identified phylogenetic analysis as abductive, that is, neither exactly deductive nor exactly inductive. This is explained by the fact that the phylogenetic inference is retrodictive analysis of historically unique events. Among other things, this means that inference from effects to cause(s) is ampliative. As systematists, we can only observe the effects under the causal theory of descent with modification (i.e., observable synapomorphies), but possible causes for conflicting data are also reversals and parallelisms. Thus, multiple conclusions about the cause(s) are possible in phylogenetic inference. By contrast, true deduction is nonampliative, enabling inferences from cause to effect(s), with singular conclusions for any given analysis.

### 5.6.2 Statistical consistency

A method is said to be statistically consistent if it converges on the true tree as progressively more character data is added to the analysis. One reason likelihood has gained popularity in the last two decades is that it has been suggested that other methods of phylogenetic inference, namely Hennigian parsimony, are statistically inconsistent under certain, largely hypothetical models of evolution (Felsenstein 1978; Penny et al. 1992). The region of statistical inconsistency has been referred to as the “Felsenstein Zone,” and it is the result of a process termed “long-branch attraction.” The long-branch attraction problem occurs when convergent homoplastic changes are more frequent than nonreversed changes in an informative part of the tree (Felsenstein 1978). This confounds Hennigian phylogenetic systematics because, under the Auxiliary Principle, the convergent homoplasies will tend to be considered as homologies and thus the taxa with their convergent “long-branches” will be grouped together (Hennig 1966; Felsenstein 1978). In simplest terms, when the data are lying about the relationships of the taxa, phylogenetic systematics (i.e., Hennig 1966) may fail to discover the true relationships. How often this occurs in nature is unknown, but Huelsenbeck (1997) has cited one case involving insects as exemplifying the long-branch attraction problem in a real data set (but see Siddall and Kluge 1997 pp 319–320). The concern that “noise,” or random data, misdirects phylogenetic systematics is also overblown (Wenzel and Siddall 1999).

Maximum likelihood has been reported to exhibit the favorable property of statistical consistency in the face of these situations (Felsenstein 1978; Penny et al. 1992; Yang 1994). It is true that in statistics the maximum likelihood estimate of a parameter is consistent (Fisher 1922; Edwards 1972). Simulation experiments
have shown this to be true in the phylogenetic context (Yang 1996), but only when the same random model used to generate the data is used and/or only when a certain correction factor is implemented (Steel et al. 1993; Siddall and Kluge 1997; Steel and Penny 2000; Goloboff 2003). It is an important caveat that maximum likelihood methods are only consistent (i.e., converge on the “true tree”) under a certain set of circumstances which typically requires that the “correct” model is used, but that the correct model and the true tree are both unknowable for real systems.

Farris (1973) suggested a protocol by which parsimony methods could be interpreted as derivatives of statistical estimation methods. This probabilistic view of parsimony was critiqued by Felsenstein (1973, 1978, 1981a, b, 1983), who focused on the statistical deficits of parsimony when viewed as a likelihood method. In general, parsimony and likelihood approaches produce the same results under the assumption of particular parameters for parsimony, i.e., low rates of evolutionary changes or equal rates of evolution among the observed lineages, or low rates of homoplasy (Felsenstein 1983). It has also been demonstrated that parsimony-based methods can be consistent (Steel et al. 1993; Siddall and Kluge 1997; Steel and Penny 2000, with a correction factor). Siddall and Kluge (1997) and Steel and Penny (2000) correctly point out that all methods are potentially consistent or inconsistent (Steel et al. 1993). Statistical consistency is not a property of a method, but the property of a specific data set, the model, and specific situations.

Likelihood methods generally require that the processes maintaining current systems were always present and consistent in the past, and this is likely not the case. Models range from relatively simple to complex, but all make approximations for computational efficiency that impair evolutionary realism. When the model is insufficient or inappropriate, as is common, arguments of statistical consistency are rendered moot. As noted earlier, frequency probability theory is violated in the calculus of current maximum likelihood methods (due to violation of independence requirements). Maximum likelihood uses a method intended to predict the future out of a range of simultaneously possible outcomes to infer a nonprobabilistic historical singularity, the evolutionary relationship of organisms.

It appears that neither epistemological nor ontological parsimony methods for inferring phylogenies can be justified, or given precedence over the other, by any part of Popper’s philosophy. Nor can either approach be given priority by appeals to statistical consistency. And, finally, each approach can be given a Bayesian interpretation. Is there an objective way to choose between, or to reconcile these subjectively divergent approaches?

An appeal to collegial pluralism (Faith and Trueman 2001) seems like a good idea at first glance. It is becoming a common practice for an author to present
maximum parsimony, maximum likelihood, and Bayesian analyses of the same data, then either arbitrarily to express a preference for one of them, or to present a consensus tree of the outcomes of each analysis, and using that as “the phylogeny.” We do not believe this is a useful approach. In the first case, the arbitrarily chosen result inevitably is the one that best supports the evolutionary scenario advocated by the author, which actually weakens the author’s case that the result emerged uniquely from the data. In the second case, a consensus tree effectively hides precisely the parts of analyses that are in need of additional scrutiny, giving author and audience a false sense of security about the results.

We believe that individual data analyses without reference an explicit evolutionary model or hypothesis (i.e., epistemological parsimony) are not explanations. They are descriptions, admittedly highly sophisticated descriptions, but just that. Fitting data to a model provides explanations, but does not give any means of testing the veracity of the model or its assumptions. Independent description and assessment relative to explanatory models both appear to be necessary but not sufficient for robust explanations. Or,

- If models do not agree with the empirical data, chances are the models, not the data, should be reevaluated. This is not an antimodel stance. A mutually reinforcing and mutually modifying dialogue between models and empirical discovery enhances progress. Brooks and McLennan (2002)

Kluge (1989, 1991, 1997, 1998a, b, 1999) has argued persuasively that historical sciences progress through cycles of discovery and evaluation, both of which are necessary but neither of which is sufficient for complete understanding, and both of which require objective methods of study. We believe that Hennigian (epistemological parsimony) analysis is the best discovery method we have in phylogenetics. This is because its results are dependent on a minimum of a priori assumptions and thus the range of potential discoveries indicated by the data is greater than for any ontological parsimony approach. At the same time, we believe that this feature of Hennigian phylogenetics renders it relatively weak as an instrument of evaluation.

It appears to us that the various maximum likelihood and Bayesian likelihood approaches are admirably suited as evaluation methods. We would like to see epistemological and ontological parsimony methods used together in a form of reciprocal illumination, not in the narrow sense of deriving a tree from multiple characters, but in a broader sense of cycles of discovery and evaluation. How would this work?

To begin with, suppose a Hennigian analysis, a maximum likelihood analysis, and a Bayesian analysis produce the same result? We should all celebrate because we would have a relatively independent discovery (the Hennigian
tree) supporting an evolutionary model (the likelihood tree). In this case, no one should have any concerns using the likelihood model to infer divergence rates on the Hennigian tree. Now, what does it mean if these different analyses do not produce the same result? The findings by Rokas et al. (2003) suggest that all such situations are an indication that not enough data have been used. Their total-genome (107 gene sequences) analysis of a clade of 7 yeast species showed that a minimum of 20 genes (i.e., roughly 20% of the genic characters possible) was required to ensure that all methods produced the same phylogenetic tree. Different methods gave different results when not enough characters were used to compensate for ambiguities of various kinds in the data, which are handled in different ways by the different methods discussed in this chapter. These results were anticipated experimentally by Hillis et al. (1992, 1993; Huelsenbeck and Hillis 1993), who produced a phylogeny for bacteriophages maintained in the laboratory. They discovered that most model-based approaches and Hennigian phylogenetics converged on the same, and true (since it was known) phylogeny as more and more traits were sampled. These results would seem to suggest that the primary response to any situation in which the different approaches to phylogenetics produce different answers should be


What do we do while we are waiting for enough data to give the same answer with all methods? Hillis et al. (1992, 1993; Huelsenbeck and Hillis 1993) showed that when a limited amount of data are available, some models generate the correct phylogeny better than Hennigian approaches. Although this finding has been interpreted by some as an indication that model-based methods are inherently superior to Hennigian methods, Hillis et al. (1994) pointed out a significant trade-off. Model-based approaches provide a distinct answer based on little data, but the confidence you have in that answer is proportional to your belief that the model used accurately reflects the evolutionary process over extended periods of time for the clade you are analyzing.

The issue becomes–how do we know the model typically gives the truth? Hillis et al. (1992) took a critical first step by generating an experimental phylogeny. The next step is to ask how typical of evolution is that phylogeny? Remember that the phylogeny (1) involved bacteriophages and (2) was generated in the laboratory according to rules invoked by the researchers (reminiscent of the Caminalcules to develop one of the first epistemological parsimony algorithms). Some have suggested that prokaryote evolution has produced not a phylogenetic tree but a highly reticulated network (Doolittle 2000), in which case the experimental phylogeny produced by Hillis et al. (1992) is not typical of evolutionary
history for their model organisms. Nonetheless, their results may still be typical of phylogenesis for eukaryotes.

More important is the question, how large a role has the historical contingency that is such a critical part of Darwinian mechanisms played in phylogenesis? Some have asserted that such contingencies do not affect phylogenetic reconstructions while others have asserted the opposite (see Yang and Bielawski 2000 for a review). Seen in this light, is it possible that the reason Hennigian and model-based approaches converge with increasing data is that the more data we consider, the more historical contingencies will play a role, in which case model-based approaches will progressively choose models whose set of “allowed possibilities” most closely approximates the minimal “a priori restrictions” of Hennigian phylogenetics. For example, Gissi et al. (2000) reported lineage-specific evolutionary rates for different mammalian mtDNA genes, suggesting that recovering the correct phylogeny might require a different mode for each gene, and used their findings to support contentions by other molecular systematists that, given uncertainty about the true phylogeny, we cannot know which model will give the correct phylogeny, and thus we do analyses of as many genes as possible to help determine the appropriate model (Mitchell et al. 2000). For those researchers, the appropriate response to the assertion that “Nature is not as simple as our models . . . . But this should not be taken as a criticism of the use of a model” (Pagel 1992) is, “Yes, it should” (Wenzel and Carpenter 1994).

Once again, we are faced with a conundrum. We suggest that the next step involves a form of reciprocal illumination. That is, use the Hennigian result, rather than Model Test, to determine the best likelihood model (i.e., what likelihood model gives the Hennigian tree). Then use the implied divergence rates from that model to add an explanatory context to the phylogenetic description without fear of introducing an element of circularity (by using the same model to build the tree that you want to test using the tree). When repeated across a variety of clades, this approach might suggest that some models fit the data more often than others, thus shedding light on the general process of evolutionary divergence patterns and rates. Perhaps some models would be suitable so rarely that they could be excluded from consideration of such general processes.

Similarly, one could use the maximum likelihood model “chosen” by the Hennigian analysis to assess the accuracy of the phylogenetic reconstruction, e.g., by linking the inferred divergence times with biogeographic or stratigraphic data. A likelihood framework has been used to incorporate or evaluate the probability of the phylogenetic hypothesis given the stratigraphic (i.e., temporal) distribution of taxa in the fossil record (Huelsenbeck and Rannala 1997b; Wagner 1998). These analyses require additional assumptions about the completeness of the fossil record and the reliability of the stratigraphic correlations. Wagner (1998)
presented an analysis of hyaenids in which the strato-likelihood tree was 10 steps longer than the most parsimonious tree, but it was putatively more consistent with the fossil record of the taxa. If the phylogenetic tree corresponds to a model that is in stark contrast with fossil evidence, its time to get more data and reassess your tree, and also time to dig for more fossils.

5.7 Conclusions: Special concerns for paleoanthropology

This is an exciting time for paleoanthropologists to be involved in evolutionary biology. New discoveries of previously unknown primates and new fossil material of poorly known taxa are constantly changing our perspectives on how our closest relatives evolved and behaved. Phylogenetic systematics is a powerful investigative tool for generating and evaluating new evolutionary hypotheses. Its power lies in its transparency and paucity of assumptions; rather than appealing to the authority of what we “already know,” parsimony analysis generates hypotheses allowing the raw data speak for themselves.

Phylogenetic systematic methodology has become widely accepted among anthropologists and paleoanthropologists in particular; most generally acknowledge its applicability to evolutionary studies of the primates (Groves 1978; Strasser and Delson 1987; Disotell 1996; Begun et al. 1997; Rae 1997; Ross et al. 1998; Harris 2000; Begun 2001; Newman et al. 2004). Many of these studies have revealed aspects of primate relationships that were previously unknown and have pointed to new avenues for research. For example, Begun’s (2001; Begun et al. 1997) analyses of 14 taxa and 240 and 247 characters, respectively, suggests that Eurasian late Miocene hominoids (Dryopithecus and Ouranopithecus) are the sister taxa to the African ape–human clade. We can use this hypothesis to draw inferences about Miocene hominoid biogeography; Begun mapped the areas in which hominoids are found onto the cladogram to demonstrate that the most parsimonious optimization is of a Eurasian origin of the African ape–human clade. Likewise, it is possible to map characters onto the tree to explore how certain characters evolve. It should be clear that to avoid circularity, the tree must be robust enough to maintain its topology without the characters under investigation (Brooks and McLennan 2002).

Phylogenetics is not without detractors in the anthropological community. Some have suggested that complicating factors specific to hominids may result in fallacious results using phylogenetic systematics (Andrews 1992; Lieberman 1999; Collard and Wood 2001; Hawks 2004). In a recently published article, Hawks explored the degree to which cladistic methodology is able to clarify
relationships within hominids. He cited difficulty with character choice, character independence, atomizing continuous traits, and coding polymorphic characters as key causes undermining the reliability of phylogenetic methodology. These factors are not restricted to hominid systematics; they constitute some of the most important and subjective decisions in phylogenetic research. There is no real consensus on how to code morphological characters, or even what constitutes a good character. A character is merely an observed part or attribute of an organism (Colless 1985), assumed to be heritable. The nature of characters is clade-dependent, and the usefulness of any particular character in a phylogenetic analysis will be directly related to whether it is a synapomorphy or not. Of course, we do not know whether a character is synapomorphic until we run the analysis, based on as large a number of characters as possible.

Concerns about character choice are based on the notion that a set of characters that have all evolved in concert, yet in parallel, as a result of convergent selection vectors, could produce a robust Hennigian result that was nevertheless an incorrect depiction of phylogeny. Collard and Wood (2001) suggested that correlated functional convergences in catarrhine face and jaws made phylogenetic analyses using only those characters unreliable. This is a problem for paleoanthropology, in which the majority of fossil remains are exclusively, or nearly so, crania and teeth (Strait et al. 1997). It is true that for any data set in which correlated homoplasies outnumber homologies, Hennigian methods will produce an incorrect result. In the Hennigian system, however, homoplasies are not known a priori, so correlated homoplasies cannot be known a priori. Therefore, we must assume character independence alongside the presumption of homology. This has recently been dubbed Kluge’s Auxiliary Principle (always presume character independence in the absence of evidence to the contrary) (Brooks and McLennan 2002). This principle asserts that the identification of potential cases of correlated homoplasies begins with a phylogenetics analysis, which identifies correlated traits. Once historically correlated traits have been identified, a researcher can ask whether or not they have unduly influenced the results. If reducing the number of phylogenetically correlated traits that come from a single part of the body significantly alters the phylogenetics hypothesis, only then is it reasonable to discuss the possibility of correlated homoplasies affecting our phylogenetics reconstructions. Similar reasoning applies to the coding of continuous (meristic) and polymorphic traits. Both pose theoretical and methodological problems for the interpretation and optimization of characters on the tree. For example, the “polymorphism” setting in PAUP* does not allow hypothetical ancestors to possess the polymorphic condition (Swofford 1998). This is a default to maintain tree length at a minimum. PAUP* treats polymorphic taxa as
heterogeneous, however some programs treat taxa possessing multiple states as missing data (Wiens 1995).

Finally, we must remember that phylogenetic analysis of any kind is merely an analytical tool. If the results of a Hennigian analysis are not what we expected, there is nothing in the methodology that requires that we believe that the results of any given analysis are true. Rather, we are faced with several possibilities: (1) the data are lying to us and no amount of additional data will rectify the problem, (2) the method is flawed, or (3) we do not have enough data. If we opt for the first two possibilities, we must then choose a model that better fits our beliefs, experience, or understanding of the evolutionary process than do the data, then reconcile the data to the model.

If we opt for the third possibility, we need to be exhaustive and explicit in character choice, to include new data and reevaluate our hypotheses frequently, and to use independent data to corroborate or refute those hypotheses. Phylogeneticists must be explicit about the methods they use to generate and evaluate trees, and those evaluating their publications need to be aware and informed of the methods used. A list of character descriptions as well as the data matrix used to generate the tree should be available so that other researchers can evaluate the authors’ character codes. Consistency indices, decay analyses, and consensus trees are important procedures for assessing the robustness of results, and should be published. As well, the techniques employed to generate these results should be clear. Other elements to consider when generating or evaluating phylogenetic hypotheses include: Do independent data (e.g., fossil dates, biogeography, behavior) corroborate or refute the hypothesis? Were all known taxa included? If not, is it because of “poor preservation, or lack of adequate sample size”?

This transparency will facilitate discussion and evaluation of various characters, and allow us to test hypotheses that, for instance, craniodental characters exhibit more correlated homoplasy than postcranial characters (Collard and Wood 2001). It will lead to discussions of functional correlations of characters, and eventual reevaluation of character choice and coding, strengthening the phylogenetic hypotheses we create. While the goal of a phylogenetic analysis is to reconstruct relationships based on shared, derived characters, the really interesting evolutionary enigmas in any clade are the homoplasies. These convergences and parallelisms direct our attention to possible functional or environmental selection responsible for the visible pattern of evolutionary change. Or,

▷ [there are] . . . four guiding principles: (1) phylogenetic trees are necessary but rarely sufficient for explaining evolutionary origins and diversification; (2) we must always be responsible for well-formulated questions; (3) we must always
be responsible for the quality of the data used in any level of our analyses, from generating phylogenetic hypotheses to testing general theories; and (4) everything we learn implies yet more cycles of discovery and evaluation . . . Brooks and McLennan (2002)

Acknowledgments

We thank the editors for inviting us to contribute to this volume. For their participation in a graduate seminar course entitled “Foundations of Phylogenetics,” we thank Jaret Bilewitch, Charmaine Condy, Nikolas Gour, Dominik Halas, Stephanie Hill, Michelle Mattern, Robert Murphy, Richard Winterbottom, David Zamparo, and David Zanatta. DCE, JF, KEF, and LAT thank Robert Reisz and Johannes Müller for thought-provoking discussions. DRB, DCE, and KEF acknowledge funding support from the Natural Sciences and Engineering Research Council (NSERC) of Canada. JF acknowledges support from the Government of Canada Awards.

References

Brooks DR, O’Grady RT, Wiley EO (1986) A measure of the information content of
phylogenetic trees, and its use as an optimality criterion. Syst Zool 35: 571–581
Felsenstein J (1973) Maximum likelihood and minimum steps methods for estimating
evolutionary trees from data on discrete characters. Syst Zool 22: 240–249
Felsenstein J (1978) Cases in which parsimony or compatibility methods will be positively misleading. Syst Zool 27: 401–410
Jaynes ET (1957a) Information theory and statistical mechanics I. Phys Rev 106: 620
Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equation of
More taxa or more characters revisited: Combining data from nuclear protein coding
genes for phylogenetic analyses of Noctuoidea (Insecta: Lepidoptera). Syst Biol 49:
202–224
Mitchell PC (1901) On the intestinal tract of
Mitchell PC (1905) On the intestinal tract
of mammals. Trans Zool Soc Lond 17:
437–536
Nelson CH, Van Horne GS (1975) A new
simplified method for constructing Wagner
networks and the cladistics of pentachaeta
(Compositae, Asteraeae). Brittonia 27:
362–372
drial phylogeny and systematics of baboons
Newton MA, Zhang YL (1999) A recursive al-
gorithm for nonparametric analysis with
missing data. Biometrika 86: 15–26
Neyman J (1974) Molecular studies: A source of
novel statistical problems. In: Gupta SS,
Yackel J (eds) Statistical decision theory and
related topics. Academic Press, New York,
pp 1–27
Nylander JA, Ronquist F, Huelsenbeck JP,
Nieves-Aldrey JL (2004) Bayesian phyloge-
etic analysis of combined data. Syst Biol
53: 47–67
estimation of ancestral character states on
Pagel MD (1992) A method for the analysis of
Penny D, Hendy MD, Steel MA (1992) Progress
with methods for constructing evolutionary
trees. Trends Ecol Evol 7: 73–79
Platnick NI, Funk VA (eds) (1983) Advances in
cladistics, vol 2. Columbia University Press,
New York
Popper KR (1968) The logic of scientific dis-
covery. Harper and Row, New York
Popper KR (1997) The demarcation between
science and metaphysics. In: Schilpp PA
(ed) The philosophy of Rudolph Carnap.
Open Court, La Salle, pp 183–226
Posada D, Crandall KA (1998) ModelTest: Test-
ing the model of DNA substitution. Bioin-
fomatics 14: 817–818
Prim RC (1957) Shortest connection networks
and some generalizations. Bell Syst Tech J 36:
1389–1401
Rae T (1997) The early evolution of the homi-
noid face. In: Begun DR, Ward CV, Rose MD
(eds) Function, phylogeny and fossils: Mioc-
cene hominoid evolution and adaptations.
Plenum Press, New York, pp 59–77
Rannala B, Yang Z (1996) Probability distribu-
tion of molecular evolutionary trees: A new
43: 304–311
Biol 52: 259–271
Rodriguez F, Oliver JL, Marin A, Medina JR
(1990) The general stochastic model of nu-
cleotide substitution. J Theor Biol 142:
485–501
Rokas A, Williams BL, King N, Carroll SB
(2003) Genome-scale approaches to resolv-
ing incongruence in molecular phylogenies.
Nature 425: 796–804
Ronquist F (2004) Bayesian inference of charc-
analysis of anthropoid relationships. J Hum
Evol 35: 221–306
Shannon C (1948) A mathematical theory of
Siddall ME, Kluge AG (1997) Probabilism
and phylogenetic inference. Cladistics 13:
313–336
Sneath PHA, Sokal RR (1973) Numerical tax-
onomy: The principles and practice of nu-
cerical classification. W.H. Freeman, San
Francisco
Sober E (1983) Parsimony in systematics: Phil-
osophical issues. Ann Rev Ecol Syst 14:
335–357
Sober E (1988) Reconstructing the past. Parsi-
mony, evolution, and inference. MIT Press,
Cambridge
Sokal RR, Sneath PHA (1963) Numerical tax-
onomy. W.H. Freeman, San Francisco
Sokal RR, Rohlf FJ (1981) Taxonomic congru-
ence in the Leptopodomorpha re-examined.
Syst Zool 30: 309–325
Sorenson MD (1999) TreeRot. Boston University, Boston
Tillyard RJ (1921) A new classification of the order Perlaria. Can Entomol 53: 35–43
6 Homology: A Philosophical and Biological Perspective

Olivier Rieppel

Abstract

Current discussions of the concept of homology center on metaphysical and epistemological issues: what are homologs, and how can they be discovered? Homology is the relation of common ancestry shared by homologs. Homologs are the “same” parts of two or more organisms under every variation of form or function that are shared due to inheritance from a common ancestor. Homology is the key to phylogeny reconstruction, because the hierarchy of homologs/homologies is isomorphic with the phylogenetic hierarchy of taxa. Consequently, the discussion of homologs as evolving entities shares many parallels with the discussion of species evolution. Homologs are conceptualized as tokens of natural kinds that engage in causal processes of ontogeny, phylogeny, and physiology. Operational criteria for the discovery of homology are the topological relations (connectivity) shared by homologs. These operational criteria are (at least to some degree) grounded in ontogeny, where developmental modules may share the relation of homology. In that sense, the operational criteria for the discovery of homology are aligned with the structure of the causal processes in which the homologs qua tokens of natural kinds engage. The discussion of concepts of partial homology and the complementarity of static (taxic, hierarchical) versus dynamic (developmental, transformational) approaches to homology reveals empirical as well as a semantic components.

6.1 Introduction

The pretheoretical use of the concept of homology reaches far back into the preevolutionary history of comparative biology (Panchen 1994). Belon (1555) is usually taken as locus classicus for the earliest illustration of the relation of homology (Figure 6.1), although the concept must have been implicit in Aristotle’s recognition of dolphins as mammals (Russell 1982 p 6).
Figure 6.1
In a comparison of the skeleton of man and a bird, Belon (1555) recognized the topological equivalence of constitutive elements (courtesy of the Field Museum Library, Mary W. Runnells Rare Book Room)
For Etienne Geoffroy St. Hilaire, the unity of type was based on the relation of “analogie,” a concept that captured the correspondence of relative topological relations (connectivité) of constituent elements of the organism tied together throughout ontogeny by the blood-vascular system (Rieppel 1988, 2001). Geoffroy’s “analogie” became Richard Owen’s “homology”, meaning “the same organ in different animals under every variety of form and function” (Panchen 1994 p 40; Williams 2004 p 192). Owen contrasted this concept with “analogy,” meaning “a part or organ in one animal which has the same function as another part or organ in a different animal” (Panchen 1994 p 40; Williams 2004 p 192). Panchen (1994 p 44) drew attention to the fact that, as used by Owen, homology and analogy were not necessarily mutually exclusive concepts, so a third term was required to characterize structures analogous between two or more organisms but not homologous. This was Lankester’s (1870) “homoplasy.” In fact, Lankester (1870) restricted the use of the term homology to its pre-evolutionary connotations; he called the evolutionary use of the concept that implies common ancestry “homogeny.”

At the hands of Geoffroy St. Hilaire or Owen, homology was meant to capture structural correspondence, rooted in “laws of structure” as expressed during ontogeny. The relation of homology (and analogy) was reconceptualized (Kitcher 1993 p 32) by Darwin (1859), who found that, on his theory, “unity of type is explained by unity of descent” (Darwin 1859 p 206). The term homology continued to apply to the same constituent parts of organisms that were recognized as homologs by Owen, but the meaning of the term changed to something roughly like “shared similarity due to common ancestry.” The history of the concept of homology has been variously sketched, sometimes in great detail (Panchen 1994; Williams 2004), and needs not to be retraced here. Instead, the focus of this chapter will be on issues that are central to the contemporary discussion of the concept of homology in systematics and developmental biology. How does “similarity” relate to the concept of homology, or, in other words, how can the relation of homology be rendered empirically accessible? Is there such a thing as “partial homology”? What does it mean to say, in Owen’s words, that homologs are “namesakes,” i.e., parts that “merit the same name” (Patterson 1988 p 605); or that “the same names can be given to the homologous bones in widely different animals”; (Darwin 1859 p 434). What does “sameness,” “structural identity” (Wagner 1994), or “historical identity” (Grant and Kluge 2004) mean in reference to homologs, or, in other words, what kind of names refer to homologs? Are homologs particulars (individuals), denoted by proper names, or tokens of natural kinds, denoted by general names?
6.2 Semantics of homology statements

To better understand the semantics of homology statements, it is worthwhile to first introduce some distinctions, derived from Lankester’s (1870) terminology (Table 6.1).

### Table 6.1
A comparison of Lankester’s (1870) and contemporary use of the terms homology and homoplasy. For further discussion see text

<table>
<thead>
<tr>
<th>Topology/Connectivity</th>
<th>Congruence</th>
<th>Incongruence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lankester 1870</td>
<td>Homology</td>
<td>Homogeny</td>
</tr>
<tr>
<td>Contemporary systematics</td>
<td>Primary homology</td>
<td>Secondary homology</td>
</tr>
</tbody>
</table>

Homology assessment has been recognized to be, minimally, a two-step procedure (Jardine 1969; Rieppel 1988, 2004a; dePinna 1991; see Brower and Schawaroch 1996 for a more elaborate account). An initial conjecture of homology, based on some similarity relations that obtain between the organisms under analysis, represents an observational statement. The relation expressed by that statement was called homology by Lankester (1870), which corresponds to the “topographic homology” of Jardine (1969 p 357) or the “primary homology” of dePinna (1991). Statements of primary homology are observational statements that relate to the organisms under analysis (Rieppel 2004a). The putative phylogenetic information content of those observation statements is subsequently tested on the basis of congruence (see further discussion later). If the statements pass the test (if they turn out coherent/congruent), they are called homology statements that correspond to dePinna’s (1991) “secondary homology” (phylogenetic homology sensu Jardine 1969). These statements of (secondary) homology imply “similarity due to common ancestry” and therefore are theoretical statements, i.e., theoretical (meta) statements about the observational (object-) statements (Rieppel 2004a). If coherence/congruence fails to obtain, the observation statements are said to be statements of homoplasy, which imply convergent similarity relations or chance similarity. Parallel evolution is said to result in a similarity relation between related lineages that is not inherited from their common ancestor. Consequently, congruence of observation statements (of primary homology) may theoretically be due to parallel evolution, but given the tools of phylogenetic systematics, there is no way to empirically distinguish (secondary) homology from parallelisms. From these considerations, Farris (1983) drew the agnostic (Sober 1985) conclusion that whereas incongruence undermines, congruence does not necessarily confirm phylogenetic hypotheses.
Observation statements deploy predicates that are meant to apply to the objects observed. For example, a chordate is an animal with a notochord, a “renate” is an animal with kidneys. The sentence “a chordate is an animal with a notochord” is of a subject–predicate structure, symbolized as $Fx$, where $F$ is the predicate “... is an animal with a notochord,” $x$ is the subject that is an animal with a notochord. More generally, “chordates” is the set of all animals that have a notochord, renates is the set of all animals that have kidneys. The fact that homology statements deploy predicates (Mahner and Bunge 1997; Wiesemüller et al. 2003) threatens to create an ontological inconsistency in an evolutionary context that is rooted in what philosophers call the “one-over-many” problem: “the selfsame property [e.g., to have a notochord] can be instantiated by numerically different things [e.g., rabbits and tigers]” (Swoyer 2000 p 2). The extension of a predicate is typically taken to be the set of all things to which the predicate truthfully applies: the extension of the predicate “... with kidneys” is the set of all animals with kidneys. But sets are abstract entities, not historical entities, and the semantic value of a predicate is usually taken to be a universal not a particular (but see Swoyer 2000 for further discussion). Homology statements predicate properties of organisms, and properties, in contrast to particulars (individuals such as objects or bodies) are universals (properties include the attributes or qualities or features or characteristics of things) (Swoyer 2000 p 2). Under the paradigm of nominalism, only particulars can claim existence (i.e., the occupancy of a particular space–time region), universals such as properties cannot. If we think of each term as referring to (denoting, designating) a thing (object, body), then predicates would have to refer to entities, with the consequence that universals, such as “redness” ($x$ is red), or “rabbit-hood” ($x$ is a rabbit) acquire some sort of Platonic existence as abstract entities. To avoid this conclusion, Devitt and Sterelny (1999) proposed to conceptualize predicates not as referring expressions but instead as expressions that apply to particulars (which in turn occupy a particular space–time region). For what would be particulars “when stripped of their clothing of universals” (Devitt and Sterelny 1999 p 278)? The predicate “... is red” can thus apply to this or that or yet another apple, the predicate “... is a tiger” can apply to this or that or yet another individual organism.

If the predicate “... is red” is applied to “this apple” (to which I point), then this apple is said to instantiate the property redness, but redness does not exist as a spatio-temporally restricted entity. In contrast, kidneys exist as spatio-temporally restricted entities: they can be removed or replaced by surgery. So what is the difference between “this apple is red” versus “this creature is an animal with kidneys”? One way to look at the issue (Haack 1998 p 133) is to claim that the predicate “... is red” marks out the set of all red things, and apples are (or can be)
members of that set. Similarly, the predicate “... is an animal with kidneys” marks out the kind of all animals with kidneys, i.e., the renates. The set of red things that includes apples, cars, and billiard balls is unlikely to be tied together by a unifying causal process, but the renates, which (for the sake of the argument) are here equated with vertebrates, are tied together by unifying causal processes, namely those of ontogeny and phylogeny. In other words, redness is not a property that is causally efficacious, the set of red things therefore not a natural kind. In contrast, to have kidneys is a causally relevant property that is shared by all renates, and it is in virtue of engaging in these causal processes that renates form a natural kind, as do vertebrates and also apples—red, green, or yellow—but not red things.

Under such analysis, individual tigers become tokens of a natural kind, namely parts of the sum of all tigers. “Tiger” thus becomes a name associated with a natural kind term, and the extension of that natural kind term (of the predicate “... is a tiger”) is the sum of all tigers, i.e., the natural kind tiger that is marked out by a common evolutionary origin. In other words, the common evolutionary origin is the relevant property that marks out the natural kind tiger. Each individual tiger that shares the same evolutionary origin is therefore a token of its kind. Each individual vertebrate is a token of the natural kind “renate”, or “chordate,” if the relevant property that marks out the natural kind “renate” or “chordate” is commonality of evolutionary origin.

Natural kinds come in hierarchies (Platts 1997): the individual lemon is not only a token of its proper kind but also a token of the natural kind “citrus fruit,” as well as a token of the natural kind “fruit.” The same is true for animals: the individual tiger is a token of tigers, of “mammals,” of “vertebrates.” A lemon can be squeezed, a tiger captured, a kidney removed or replaced. The sum of all kidneys constitutes a natural kind, its tokens are instantiated by all renates. “Notochord” is a natural kind term, the tokens of its kind are instantiated by all chordates. Renates and chordates constitute a natural kind, if they share a common evolutionary origin; particular kidneys and notochords (taken from rabbits or tigers) are tokens of natural kinds that share a common evolutionary origin if kidneys and notochords are homologs of vertebrates. If kidneys and notochords were homoplasies (e.g., homoplastic in rabbits and tigers), they would be tokens of nominal kinds (the corresponding terms nominal kind terms). The key to “natural kindness” is the fact that the tokens of the kind take part in causal processes, that their properties are causally efficacious (Sober 1981; Quine 1994; Shoemaker 2003).

Conceptualization of homology in terms of natural kind term semantics offers a key to why homology statements have to be tied to a conditional phrase (Bock 1974), given the fact that natural kinds, as well as homologies, come in
hierarchies. Let us follow Quine (1964) into a foreign country where people speak a language that I do not understand. In that situation, a native points at a passing rabbit and utters “gavagai.” Seeing what the native points at and hearing her utterance does not tell me whether the native means “rabbit,” “rodent,” “mammal,” or “vertebrate,” etc. In a similar sense, the statement that “bird wings and bat wings are homologous” remains underdetermined. Given that statement only, I cannot decide whether it is right or wrong relative to the background knowledge of the currently accepted theory of vertebrate phylogeny. But if the statement is “bird wings and bat wings are homologous as (qua) tetrapod limbs,” I know that the statement is right; if it is “bird wings and bat wings are homologous as (qua) wings,” I know it is false, given the background knowledge (see the discussion of the “qua‐problem” in Devitt and Sterelny 1999).

Homologs have been claimed to be individuals, or historical individuals, on ontological grounds (Grant and Kluge 2004). The reason is that only particulars, or individuals, can take part in causal processes such as descent with modification. Sets, or classes, cannot. If homology is a property, and properties are universals that mark out sets, or classes, then homologs cannot evolve. On the analysis described previously, causally efficacious properties mark out natural kinds. The common evolutionary origin marks out the natural kind tiger, but it is individual tigers, i.e., the particular tokens of the natural kind that engage in causal processes such as descent with modification. If homologous, the sum of kidneys constitutes a natural kind, but again it is individual kidneys, i.e., the particular tokens of the natural kind that engage in causal processes such as ontogeny. However, ontogeny, or descent with modification, requires plasticity if evolution is to occur. In order for this to be possible, an adequately weak notion of a natural kind must therefore be invoked such as the “homeostatic property cluster natural kind” proposed by Boyd (1991, 1999).

Common evolutionary origin can be the (essential: LaPorte 2004) property that marks out biological natural kinds, like taxa or homologs, but tigers and their parts, such as their kidneys and their parts, such as kidney tubules, share other properties as well. For example, tigers share a genetic constitution, which they transmit—more or less reliably—to offspring as they engage in the causal process reproduction. Kidneys engage in the causal process of excretion, a function that changed in some regards as renates evolved from water to land. Tokens of natural kinds which in virtue of their properties engage in developmental and physiological processes must be capable of plasticity if evolution is to occur. On the other hand, variation must not be so rampant as to disrupt ontogeny or physiological function. The clustering of properties in a variable and partially overlapping pattern that characterizes homeostatic property cluster natural kinds is believed to be effected by homeostatic mechanisms, which are
themselves plastic and capable of variation and change (West-Eberhard 2003). “Because the clustering in the world that defines such a kind can be an ongoing process, the properties in the cluster can sometimes vary over time, and the kind’s boundaries can be historically delimited if the clustering is historically delimited” (Keller et al. 2003 p 105). With genealogy itself being a homeostatic mechanism of property cluster natural kinds (Boyd 1999), the latter can also be conceived of as historical kinds (Sterelny 1994 p 11), a perspective that accommodates species as well as homologs.

6.3 Similarity, homology, and congruence

The relation of homology is a theoretical concept because it is based on common evolutionary descent. Since common evolutionary descent cannot be observed, it must be inferred from observation. Traditionally, biologists will take some sort of similarity relation to be indicative of common ancestry (Rieppel and Kearney 2002). Indeed, under the paradigm of evolution, homology is frequently defined as “similarity due to common descent” (West-Eberhard 2003 p 485).

However, similarity is a highly loaded concept. Goodman (1972 p 437) called similarity “a pretender, an imposter, a quack,” whereas Hull (1999 p 35) found it “unanalyzed, perhaps unanalyzable.” “As philosophers have long recognized, similarity without theory is empty” (Sober 1984 p 336; see also Ruse 1988 p 60; Dupré 1993 p 45). So “similarity” is taken to be a primitive concept that cannot be further defined, and yet it is hardly possible to “imagine a more familiar or fundamental notion than this, or a notion more ubiquitous in its application” (Quine 1994 p 42).

In the face of these problems, contemporary systematics has experienced the revival of an outdated empiricism (Sneath and Sokal 1973) that was thought to have been put to rest long ago. It consists in a purely instrumental–algorithmic approach to homology in which the array of all phylogenetic hypotheses possible for N number of terminal taxa is used as an instrument to generate statements of putative homology. As argued by Hárlin (1999 p 497), “too much emphasis is put on particular characters” given the “logical priority of the tree over the characters.” Similarity as a tool of systematics is declared bankrupt, and theory is eschewed on grounds of lack of positive knowledge. Critical character analysis is rejected as a “disturbing trend toward data selectivity” (O’Leary et al. 2003 p 861) by proponents of the “total evidence approach,” who cite Kluge (1997) in support of the notion that inclusion of all data, including all “published data,” results in a stronger test of homology and phylogenetic relationships (O’Leary et al. 2003 p 862). But to eschew theory in favor of an instrumentalist–algorithmic approach
to character analysis does not mean to take an atheoretical stance, and as pointed out by Ruse (1988 p 60):

- As soon as one starts breaking organisms into parts, one must bring in theory . . .
  Take two bears, one white and one brown. Do they differ in one feature, or does one take each hair separately . . . The point is whether someone who explicitly eschews . . . theory has the right to combine all the hairs into one feature.

  Proponents of the “total evidence” approach (Kluge 1989) argue, “there is no one operation for determining character states in this system—it can be anything that leads to the testable hypothesis of synapomorphy” (Kluge 2003a p 356; emphasis added). Even more drastic yet more sophisticated is Griffith’s (1999 p 225; emphasis added) assertion that

- cladistic analysis can proceed from a list of arbitrary measurements by looking for congruence among the evolutionary trees produced by different measurements . . . thus ‘bootstrapping’ itself into a reliable character set.

  Such a data set, “holding itself up by its bootstraps” Audi (2003 p 187), has been called the result of “phenetic cladistics: elegant analyses with many sources of error” (Wägele 2004 p 109). “We might just [as well] stop tugging at our bootstraps altogether” (Quine 2001 p 457).

  Contemporary systematists see homology assessment as a two-step procedure: the conjecture of homology (primary homology statements \textit{sensu} dePinna 1991), and the test of those on the basis of congruence (yielding secondary homology statements \textit{sensu} dePinna 1991). With a misguided appeal to Popper’s philosophy of science (Kluge 2003b, 2004), the proponents of the phenetic approach to cladistics consider the contextual origin of primary homology statements (conjectures of synapomorphies in Kluge 2003a) irrelevant, \textit{as long as} there is a method to test those statements. The test of choice is not Popperian, however, but built on congruence. The issue is not whether an observation statement is accepted as falsifying the theory from which its negation was deduced, as is the case in Popper’s hypothetico-deductivism (Rieppel 2003a, 2004b). Instead, the issue is whether character statements cohere to a degree that seems sufficient for the (provisional) acceptance of a phylogenetic hypothesis.

  Patterson (1982 p 74) considered Wilson’s (1965) consistency test for phylogenies “a more rigorous formulation of my congruence testing.” The work done by the test of congruence is to maximize the consistency of all available data in marking out an enaptic (inclusive) hierarchy. In a “pattern cladist” approach, Patterson (1982) took characters to mark out sets and analyzed the consistency of their distribution in terms of mutual inclusion, mutual exclusion, or overlap of the sets the character statements define. The same logic underlies “Hennig
argumentation” (Hennig 1966), ultimately based on the “inclusion/exclusion principle” (Wiley et al. 1991 p 45). Accordingly, the test of congruence is based on logic as it tests for coherence versus contradiction of character statements, but logic has a linguistic foundation (Soames 2003 p 136). Coherence and contradiction are logical relations that obtain between sentences only and the propositions they express, not between organisms and their parts. Kidneys and notochords cannot cohere or contradict each other, only statements about kidneys or notochords can. The test of congruence therefore does not test for descriptive accuracy, causal relevance of character statements, or the like. It only tests for coherence (consistency relative to a hierarchy of sets within sets, or, in the present analysis, kinds within kinds) of character statements. That is why phenetic cladistics is prone to many sources of error. If the contextual origin of conjectures of homology (primary statements of homology) were indeed irrelevant, and if the test of congruence were indeed sufficient for phylogeny reconstruction, there would be no theoretical basis on which to reject the purely operational “punch-card approach” (Sneath and Sokal 1973, Figure 3.1) to the automatic scanning of characters. If congruence and nothing else is used in a test of character statements, then the support for a phylogenetic hypothesis derives exclusively from the mutual support of the coherent character statements, without consideration of how those character statements do or do not refer (apply) to the organisms under analysis. The individual character statements, and with them the phylogenetic hypothesis they putatively support, potentially remain hanging in the air. Haack (1998 p 27) found such inference to fall victim to the “drunken sailor argument” that invokes the picture of two drunken sailors who believe they can mutually support each other by standing back to back when both stand on nothing.

Of course there can be no doubt that character congruence, i.e., the coherence of character statements relative to a hierarchy, is a necessary condition for phylogeny reconstruction. Incoherent talk about the world makes no sense, not in phylogenetics, nor in any other domain of discourse. But the mere coherence of character statements is not also a sufficient condition for phylogeny reconstruction. If they are not to be left hanging in the air, character statements must somehow be grounded in the organisms under analysis, i.e., rooted in causal theories of inheritance, development, and/or function. Homologies are natural kinds only if their tokens (at least potentially) take part in causal processes. To thoroughly investigate the genetic, developmental, even functional background of each character used in phylogeny reconstruction is beyond practical reach, however. Homology is a theoretical relation that has to be inferred from observation, but observation cannot proceed meaningfully without theory and outside the context of background knowledge (Hanson 1961; Haack 1998). Quite simply
put: to be relevant to a theory, the predicate deployed by an observation statement must have a theoretical counterpart in the theory (a set of theoretical sentences) to which the observation statement is meant to relate to (Körner 1970; Mahner and Bunge 1997). Theory and background knowledge will therefore have to be brought to bear on observations that result in the formulation of primary homology statements. Even if the distinction is unsharp and sometimes difficult to draw, this is how relevant evidence is generally distinguished from total evidence (Haack 1998). More specifically, a method is required that will allow to ground primary homology statements at least approximately, and certainly defeasibly (the hypothesis of primary homology may be wrong) in causal theories of inheritance and development. This method consists in the application of the “operational criteria of homology,” such as topology and connectivity (Remane 1952; Riedl 1977 showed that the additional operational criteria of homology formulated by Remane 1952 ultimately all reduce to topology and connectivity).

Quine (1994 p 43; emphasis added) found that “there is something logically repugnant about” similarity in its primitive, innate, unanalyzed, and unanalyzable sense (Goodman 1972; Hull 1999). But science sheds this kind of similarity as it matures. Quine (1994 p 53) used chemistry as an example of a maturing branch of science, in which “comparative similarity of the sort that matters for chemistry can be stated . . . in terms of chemical composition. Molecules will be said to match if they contain atoms of the same elements in the same topological combinations.” The causal efficacy of chemical kinds is thus explained as a consequence of the topological relationships of their constituent elements. This is exactly the kind of similarity that also matters for the relation of homology. Systematists do not look for any kind of similarity, but for a special kind that it is established on the basis of operational criteria of homology, i.e., the relative topological position and connectivity of constituent parts of an organism (both morphological and molecular) in their temporal (ontogenetic) and spatial (ontogenetic, molecular) manifestations. Topology and connectivity yield the special kind of similarity that delivers promissory notes for potential homology. Same-ness or (historical) “identity” has been invoked for homologs that carry the same name, but there is no “perceptual sameness” (Kluge 2003b p 236, and n. 12) in a comparison of the lower jaw of a shark and the outermost ear-ossicle of a mammal. However, there is perceptional topological correspondence (during early ontogenetic stages) that justifies the conjecture of homology (Reichert 1837).

The operational criteria of homology can be compared to “correspondence rules” (Rieppel 2003b) which, stripped of their positivist background, mean nothing more than either an experimental setup or an observational procedure
Mahner and Bunge (1997) use the broadly equivalent term “indicator hypothesis” which, in the present context, would be hypotheses that do not tell us what homology is, but instead guide us how to “look for” potential homology. The operational criteria of homology have empirically been found to be eminently successful in the reconstruction of the tree of life (Remane 1952; Riedl 1977), which is why it may be assumed that they provide a (defeasible, i.e., potentially mistaken) proxy for the causal grounding of hypotheses of homology in theories of inheritance, development, and evolution. In Boyd’s (1991, 1999) terms, the use of the operational criteria of homology seems to be well aligned with the causal structure of the world, at least to some degree, and defeasibly so (Rieppel and Kearney 2002). The use of operational criteria of homology cannot, therefore, be theory free. They do not establish a purely “operational” link that maps a theoretical statement of homology on to an observational statement of topological correspondence of parts. The theoretical background that justifies the use of operational criteria of homology is the recognition that topology and connectivity are expressed in development, and that evolutionary transformation of morphology occurs through the transformation of development (see further discussion later). A similar adherence to correspondence rules is known in molecular systematics: “Alignment of sequences requires explicit and objective rules if inferences of positional homology are to be robust” (Hillis 1994 p 349; but see Wheeler 1998).

6.4 Development and homology

Homologs share a common evolutionary origin, but “neither traits nor representations of traits are transmitted to offspring. Instead, they are made—reconstructed—in development” (Oyama et al. 2001 p 2). Both development and evolution can be conceived of as a “process of construction and reconstruction in which heterogeneous resources are contingently but more or less reliably reassembled for each life cycle” (Oyama et al. 2001 p 1). Development results in pattern formation, an “activity by which embryonic cells form [spatially] ordered arrangements of differentiated tissues” (Gilbert 1988 p 589), and comparative morphology analyzes these patterns in terms of topology and connectivity. Raff and Sly (2000 p 102) noted that developmental modules “occupy specific physical sites within the embryo,” and “exhibit varying degrees of connectivity to other modules within the embryo.” This is the juncture at which the operational criteria of homology find their partial, and defeasible, causal grounding in the process of development. It is for these reasons that developmental modules, themselves natural kinds (Wagner 1996, 2001; Brandon 1999; Rieppel 2004c), have been
identified as the “traditional units for claims about homology” (Winther 2001 p 117). However, the use of ontogeny to causally ground homology statements is riddled with difficulties.

For Darwin (1859 p 449), “… community in embryonic structure reveals community of descent. It will reveal this community of descent, however much the structure of the adult may have been modified and obscured.” However, in the fourth edition of the Origin (1866 p 312; see Peckham 1959), Darwin added: “… community in embryonic structure reveals community of descent; but dissimilarity in embryonic development does not prove discontinuity of descent.” The reason is that structures considered homologs may share strikingly different developmental trajectories. A simple and classic example is the generally accepted homology of the vertebrate brain in spite of the fact that the neural plate delaminates, rather than invaginates, in teleosts. Hall (1995) presented a review of the role of ontogeny as a criterion of homology, concluding that “homology can no longer retain its historical links to shared embryonic development,” for “there are so many examples of homologous structures arising from nonhomologous developmental processes” (Hall 1995 p 8; see also Rieppel 1993; Hall 1994; Rieppel and Kearney 2002). But just as potential homologs can develop along different ontogenetic trajectories, so can a similar developmental background generate very distinct morphologies: “The transcription factors ‘distal-less,’ ‘engrailed,’ and ‘orthodenticle’ each have orthologs involved in patterning very different structural features in different metazoan taxa” (Mindell and Meyer 2001 p 435). Similarly, Raff (1996 p 352) found it surprising “that both genes and principles of appendage development are shared between two phyla,” i.e., arthropods (Drosophila) and vertebrates (tetrapods). As summarized by Wagner (1994 p 275), organs with the same structural organization (structurally identical body parts) may have different developmental pathways, may use different genetic information for their development, and may not share a common phylogenetic origin.

Developmental biologists therefore recognize different concepts of homology, of which the one discussed in the preceding section—the phylogenetic or “taxic” concept of homology—is only one. In addition, there is the “morphological homology concept” that uses “structural identity for its definition,” and the “biological homology concept,” which is based on “shared developmental constraints” (Wagner 1994 p 174; Roth 1984 emphasized shared developmental pathways instead). According to the latter concept, “structures from two individuals or from the same individual are homologous if they share a set of developmental constraints, caused by locally acting self-regulatory mechanisms of organ differentiation. These structures are thus developmentally individualized parts of the phenotype” (Wagner 1994 p 275; see also Wagner 1989).
In contrast, West-Eberhard (2003 p 486) distinguished what she called “broad-sense homology” (corresponding to “biological homology”) from “cladistic homology.” In contrast to cladistic homology, broad-sense homology allows for parallelisms (see the discussion above on the problem of an empirical distinction of parallelisms from taxic homology), and the recurrence of features previously lost in phylogeny. As such, broad-sense homology includes cladistic homology as a subset, the two concepts in fact representing “complementary approaches to understanding the evolution and stability of structure” (West-Eberhard 2003 p 488; on the issue of complementarity see the discussion later). Indeed, statements of biological homology and statements of phylogenetic (taxic, cladistic) homology may be expected to be congruent if the developmental mechanisms underlying the corresponding structures are so complex that their repeated evolution would seem very unlikely (Shubin 1994 p 206). However, because of the problems outlined previously, “there is no rigorous way in which biological homology concepts can be used to distinguish between phylogenetic homology and homoplasy” (Shubin 1994 p 206). It is for these reasons that comparative studies of development “can be used to trace homology, but developmental differences do not negate it” (West-Eberhard 2003 p 496). Similarity of gene expression patterns and developmental pathways as captured by biological homology may support hypotheses of taxic homology but not the other way around (Rieppel 2004c).

6.5 Partial homology or partial reference?

If true that development is a “process of construction and reconstruction in which heterogeneous resources are contingently but more or less reliably reassembled for each life cycle” (Oyama et al. 2001 p 1), a situation might arise in which a developmental biologist would want to invoke a concept of partial homology. The concept of modular development (Raff 1996; West-Eberhard 2003; for a review and references see Rieppel 2004c) builds on connectivity yet dissociability of developmental modules. If developmental modules are dissociable, they can potentially “be reorganized somewhat independently of each other” (West-Eberhard 2003 p 60), thus providing the resources for reassembly of traits derived from heterogenous resources in descendant life cycles. Such “mosaic development” results in heterochrony, heterotopy, and other effects (Raff 1996) that led West-Eberhard (2003 p 60) to speak of homology as “not an all-or-nothing phenomenon but rather a continuum of more or less similarity due to common ancestry.” This conclusion, startling from a systematist’s point of view, would
seem to be supported by Raff’s (1996) review of appendage development in *Drosophila* and tetrapods. Why not speak of partial homology of appendages in those widely separated lineages in which “appendages have a deep underlying and (prior to the findings of developmental genetics) completely unexpected homology” (Raff 1996 p 353; see also Shubin et al. 1997; Gilbert and Bolker 2001)? Very basal building blocks are used to develop vastly different structures. But is it true that through such mosaic development, “evolution makes a mess of homology” (West-Eberhard 2003 p 497)? Such a conclusion seems subject to a naturalistic fallacy. Homology is a theoretical concept. It may, indeed, be the case that our theoretical concepts are somewhat messy, but this is not because of evolution. The evolutionary process just unfolds the way it does, and this may be a way that is difficult to capture conceptually.

The call for partial homology can have two sources: the heterogeneous combination of ancestral resources in descendant life cycles as outlined previously, or gradual and continuous morphological change as in Sattler’s (1990, 1996, and references therein) “continuum morphology,” which essentially reduces structure(s) to process(es). Although at odds with cladistics, Weston (2000 p 141) considered it an avenue toward holism if partial homology was fitted as a distance measure on a tree. Why this appeal to holism?

The corollary of “continuum morphology” is that it “acknowledges gradations between typical structures . . . . From this point of view, homology is a matter of degree” (Sattler 1996 p 577). The tension identified by Sattler (1990, 1996) between “typical structures” and the “continuity of developmental processes” broadly corresponds to the tension between “evolution and stability of structure” that was used by West-Eberhard (2003 p 488) in support of her claim for complementarity. Indeed, the contrast between pattern and process can quite generally be conceptualized in terms of complementarity (Rieppel 1988), although this may in turn result in some tension with scientific realism that adheres to ontological monism. Ontological monism holds that there is one, and only one, way to carve nature at its joints, i.e., that particles and waves are not two different kinds of light but rather a reflection of our imperfect understanding of light. Accordingly, McMullin (2002 p 252) argued that complementarity invoked in physics does not also invoke an ontological dualism, but rather points to the issue of partial reference of the terms deployed in complementary theories. The same seems true for biology.

The discussion of partial homology in terms of morphological stasis versus continuity of developmental change mirrors the paradoxical discussion of “species” as spatio-temporally delimited chunks of a genealogical continuum. With respect to species, Hey (2001 p 168) found that “systematists and
evolutionary biologists have both placed incommensurable demands on a word,” in which the word “species” is supposed to function in the context of pattern as well as process analysis (Hull 1997), and the same is true of homology. But again: “a tendency for different perspectives by different biologists [must not] be confused with different kinds of reality” (Hey 2001 p 150). Indeed, if species transformation is a gradual process, any species concept will be subject to the “Sorites Paradox” (Rieppel 2003b): take a heap of sand and remove grains one by one—when would one stop calling it a “heap”? Under the paradigm of gradualism, species boundaries will be unsharp, fuzzy, no matter which concept is deployed in their delineation (Hey 2001). The same is true of homology in cases in which morphological transformation (developmental and phylogenetic) is a gradual process. There is, however, a conceptualization of species that specifically allows for such fuzziness, which is that of homeostatic property cluster natural kinds (Boyd 1999; see also Griffiths 1999; Wilson 1999; Keller et al. 2003). The same concept can be applied to homologs, as was discussed previously (Rieppel 2004c). The conclusion is not that evolution makes a mess of homology, but that it is the names associated with the natural kind terms (predicates) we deploy in statements of homology that may have imprecise, or even only partial, reference (the terms apply only imprecisely, or partially, to the organisms under analysis).

A second issue to consider in relation to the concept of partial homology is the fact that natural kinds, such as taxa, or homologs, come in hierarchies and thus raise the qua-problem for natural kind terms as discussed previously (Devitt and Sterelny 1999). The statement “bird wings and bat wings are homologous” makes little sense unless it is tied to a conditional phrase. The same is true of the statement “arthropod and tetrapod appendages are partially homologous,” unless the question is raised in which sense they are homologous, partially or not. The concept of partial homology may reflect a tendency of developmental biologists to use broad-sense homology independent of a hierarchical conception of development. If the sense of partial homology in this example relates to certain gene expression patterns (a “similar genetic cascade”) (Shubin et al. 1997 p 646), then the homology relation exists not at the level of appendages, but at the level of genes, which ties the homology relation to its proper hierarchical level. Gilbert and Bolker (2001 p 447) speak of such gene-level homology as “deep homology,” for insect “legs” and vertebrate “legs” obviously cannot be homologous as appendages, because “phylogenetically intermediate taxa (particularly basal chordates) do not possess comparable structures” (Shubin et al. 1997 p 645). However, Gilbert and Bolker (2001 p 449) added a note of caution to this tale of wings and legs. Applying the “small toolbox argument” to developmental processes in
the name of parsimony, they draw attention to the fact that genetic resources that regulate basal (early) patterning processes in development are limited, for which reason convergent gene expression patterns are difficult to rule out. For example, much of the “toolbox” that is involved in early limb development is also involved in the early development of vertebrate teeth.

6.6 Taxic versus the transformational approach to homology

The phylogenetic literature that deals with species, taxa, as well as with homologs, is replete with dismissals of taxa, or homologs, as natural kinds, for these are considered abstract entities, like sets, or classes (Hull 1999). The same is true of properties (attributes), which are dismissed as universals that cannot be instantiated by evolutionary entities (Grant and Kluge 2004). Hey (2001 p 153) for example dismisses property cluster kinds as polythetic classes, where “a polythetic class is just a category in which the members need only meet some, not all, of the membership criteria.” But as emphasized previously: for property cluster kinds (as well as for other kinds) to be natural requires that their properties are relevant to causal processes, i.e., that tokens of kinds take part in causal processes by virtue of their properties (Sober 1981; Boyd 1999; Shoemaker 2003; LaPorte 2004; Lipton 2004).

Homologs (Grant and Kluge 2004), as well as species (Hull 1999), have been considered to be particulars (individuals). This move was motivated by a certain reading of Hennig (1950, 1966), which found him to have replaced “intensional” with “extensional” thinking in systematics (Dupuis 1984). Such reading of Hennig has led to the assertion that systematics must rely on “ostension” and “extensionality” (Kluge 2003a). The meaning of extensionality as used in this context remains obscure (Dupuis 1984 p 13: “Hennigian taxonomy is extensional since, for appraisal of each attribute, it requires an extensive consideration of various objects”), whereas the relation of ostension is more easily understandable. In the simplest case, ostension is the relation of “pointing at” an object in order to establish that object as the referent of a proper name, or of a sentence of the subject–predicate structure (Fx). But for Grant and Kluge (2004), it is not objects and their properties that serve as homologs. Instead, “only the transformation from one character state to another, a → a’, constitutes evidence in phylogenetic inference” (Grant and Kluge 2004 p 25). If only transformation events constitute evidence, and if systematics is to be ostensional, then systematists would have to be able to “point at” such transformation events.
The collapse of a bridge or the burning of a house are events that witnesses can point at, but the transformation events that are of interest to systematists are events of the past. What is required, therefore, is identification of character states and inference of their transformation. “However, character state transformation has always been treated primarily as a metaphor … Furtherermore, character states themselves can be seen as manifestations of the transformational metaphor” (Weston 2000 p 135). In this quote, I would replace the term “metaphor” with that of “concept,” for as convincingly argued by Hanson (1961), perception itself entails a conceptual component (Rieppel 2004a). As noted earlier, the mere pointing at a rabbit and the utterance of “gavagai” by a native speaker does not tell the English speaking audience what the native speaker means by gavagai: a rabbit, an undetached rabbit-part, or a rodent (Quine 1964)? It seems impossible to meaningfully talk about the world without concepts (Luntley 1999).

A transformation series of character states purported to be homologous does not result from mere ostension but from character conceptualization (“transformation series analysis . . . begins with the construction of an initial transformation series”: Kitching et al. 1998 p 218; my emphasis). Hennig (1966 p 94) was very clear on this issue: “the belonging of characters to a phylogenetic transformation series . . . cannot be directly determined.” To solve the problem, Hennig (1966 p 94) turned to an account (in Cain and Harrison 1958 pp 86–88) of J.H. Woodger’s analysis of morphological comparison in the language of logical positivism (Rieppel 2003b), which “to that extent is exact” (Hennig 1966 p 94). According to Woodger (Cain and Harrison 1958), morphological comparison individuates parts of organisms in terms of their relation, and compares them in terms of their properties. The description of properties of, and relations between, objects are the cornerstones of the positivist philosopher Rudolf Carnap’s “construction theory,” i.e., the theory of how science reconstructs the logical structure of the world (Mayhall 2002). As argued earlier, the same theoretical background underlies Hennig’s (1957, 1966) conceptualization of the cladogram (Rieppel 2003b). There is no need to look back on Carnap, Woodger, or even Hennig, however, to understand that the identification of characters, character states, and transformation series entails a conceptual component (Hanson’s (1961) work was, after all, a stab at the positivists’ adherence to something akin to “simple” observation). In his talk announcing his own transformation, Farris (Mitter 1980 p 188; see also Dupuis 1984 p 16) asserted that “morphologists do not sample characters, they synthesize them.” If “extensional” systematics must be based on ostension, but if it admits only character transformation as evidence, then the argument boils down to the claim that systematists must be able to “point at”
concepts. This conclusion merely shows that such arguments confound concepts with objects (Mahner and Bunge 1997).

The discussion of transformation series reflects back on the “transformational” versus “taxic” approach to homology that was hinted at by West-Eberhard (2003) in her distinction of broad-sense versus cladistic homology. The essence of this debate (that started with Patterson 1982) is that the transformational approach allows homology to encompass both symplesiomorphy and synapomorphy, whereas the taxic approach requires homology to be synonymous with synapomorphy (it requires a statement of homology to be coextensive with a statement of synapomorphy). Wiesemüller et al. (2003 p 89) consider the taxic approach as too coarse grained because it blurs the distinction of plesiomorphy and apomorphy, in addition to the fact that the concept of homology cannot easily be applied to “absence”-characters. The latter problem was (arguably) “solved” semantically: snakes do not have “no legs,” instead they have “modified legs,” i.e., “lost legs” (Platnick 1978; see the discussion of the deduction of different meanings from different premises in Rieppel 2004b p 79). The distinction of homology from symplesiomorphy and synapomorphy on the other hand is related to the *qua*-problem. “Bird wings and bat wings are homologous” violates the requirement for coextensionality of statements of homology and of synapomorphy because it implies a symplesiomorphy: birds and bats share tetrapod (fore)limbs, not wings. Patterson’s (1982) goal simply was to remove ambiguity, and render statements of homology testable by congruence, in claiming that homology statements should always be assessed as (*qua*) statements of synapomorphy—since what is a symplesiomorphy at one level (jaws of amniotes) becomes a synapomorphy at a more inclusive level (jaws of gnathostomes). Interesting evolutionary talk about a comparison of bird wings with bat wings would be talk about convergence, but then bird wings and bat wings could not be homologous.

### 6.7 Conclusions

Homology is one of two perennially discussed topics in comparative biology—the other is species. The two discussions show many parallels, as indicated previously: proper names versus general names, individuals versus classes, sets, or kinds. To turn homologs into individuals and anatomical (or molecular) terms into proper names (Grant and Kluge 2004) is patterned on the argument that species are individuals (e.g. Hull 1999). This latter argument has received a lot of attention from evolutionary biologists, who largely missed an important
part of its theoretical foundation. In his defense of the argument from a philosophical point of view, Hull (1976 p 179, n. 4) drew attention to the semantic behavior of species names in evolutionary theory, which in his analysis corresponds to the semantic behavior of proper names as “rigid designators” (Kripke 2002). Such behavior of proper names is tied to a specific theory of reference for proper names, i.e., the “historical,” “causal,” or “direct” theory of reference (Hull 1976 p 179, n. 4; see also Hull 1988). These are all rather technical issues that need not be reviewed here; there is also no need to deliver a verdict on the ontological status of species here. The important point is that authors who want to use—in an evolutionary context—the “names” that refer to “the same organ in different animals under every variety of form and function” as proper names will need to worry about those technicalities but have not done so far.

In the present context, homology is conceptualized in terms of natural kind term semantics, and the names associated with natural kinds can be general names or singular terms (Soames 2002). Kripke (2002) himself already expanded his theories to also apply to natural kind terms, such as “tiger”, “water,” or “gold,” but as noted by Devitt and Sterelny (1999), the use of such terms — at least in the case of biological natural kinds — at some level involves some descriptive account (for Kripke 2002, a descriptive account may help to fix reference, but does not determine reference, of proper names or natural kind terms). This means that the use of names (associated with natural kind terms) to refer to “the same organ in different animals under every variety of form and function” must be tied to a conditional phrase that specifies the hierarchical level at which the use of that name gains some cash value in terms of marking out monophyletic groups. This in turn means that phylogeny reconstruction cannot be a matter of mere extensionality and ostension (Kluge 2003a, b; Grant and Kluge 2004), but requires a conceptual, i.e., a theoretical framework. There is no immediate access to objective reality, but this only means that careful phylogenetic analysis will require at least an attempt to causally ground hypotheses of homology, proximally in criteria of topology and connectivity, ultimately in the theories of inheritance, development, and evolution.

**Acknowledgments**

The content of this paper profited much from discussions with Maureen Kearney and Richard A Richards, which is not to say that they agree with all the views expressed here. Support from the National Science Foundation is greatly acknowledged (grant # DEB-0235618, to M.K. and O.R.).
References

Quine WV (1964) Word and object. The MIT Press, Cambridge MA
Reichert C (1837) Über die Visceralbogen der Wirbelthiere im allgemeinen und deren Metamorphosen bei den Vögeln und Säugethieren. Arch Anat Physiol Wiss Med 1837: 120–222
Sneath PHA, Sokal RR (1973) Numerical taxonomy. WH Freeman, San Francisco
7 Taphonomic and Diagenetic Processes

Gisela Grupe

Abstract

The recycling of matter within an ecosystem is a fundamental process and therefore, fossilization of a body or its parts is always the exception to the rule. The transition of organic remains from the biosphere to the lithosphere (= taphonomy) comprises the successive steps of necrology, biostratinomy, burial, and diagenesis. Focusing on the taphonomy of vertebrate skeletons, fossil types, and the main processes leading to preservation and/or destruction of a dead body and how these are intertwined, are introduced. All in all, fossilization is not a random process. Almost all of the first-order changes a dead body is subject to prior to fossilization may lead to alterations in size and shape of a skeletal part, which might be mistaken for artificial manipulations (pseudoartifacts). Taphonomic processes without doubt lead to a stepwise loss of information about the formerly living being. Today, methodological progress especially in the field of archeometry permits the evaluation of a variety of lifetime parameters. However, deep insights into taphonomic, especially diagenetic, processes are the indispensable prerequisites.

7.1 Introduction

The recycling of matter is one of the fundamental processes of life. Fossils, on the other hand, are evidence for the possibility that either complete organisms or their parts may be transferred from the biosphere to the lithosphere with preservation of morphological and even biochemical features. With regard to the various biogeochemical cycles, which are characteristic for ecosystems, fossilization is necessarily the exception to the rule.

7.2 Taphonomy

Taphonomy is the paleontological subdiscipline that unravels the processes an organism is subject to from its death until its recovery (Figure 7.1). The term
“taphonomy” was introduced by the Russian paleontologist Efremov (1940) as the study of the transition of organic remains from the biosphere to the lithosphere, whereby the etymological origin of the term from the Greek taphos = tomb implies the necessity that the dead body or its parts become buried sooner or later after death (note the origin of the term “fossil” from the Latin word fossilis = excavated, related to fodere = to dig, or effodere = to excavate). Carcasses that remain on the surface usually do not have a chance to turn into a fossil and will instead undergo complete decomposition (Miller 1975; Gill-King 1997). The preserved fossil record is greatly biased toward organisms, which have durable, hard parts like mineralized tissues, and to those organisms which occur in large numbers, hence with high-population density. Fossilization therefore is a nonrandom process.

### 7.2.1 Taphonomic analysis

The first step of taphonomic analysis comprises necrology, the actual death of an organism. While most animals die in toto, entire plants frequently do not but rather shed parts like leaves or pollen, which may fossilize. After death, biostratinomy is responsible for the further condition of a fossil prior to its burial and includes processes like disarticulation (by scavengers or intended...
dismemberment by butchering), weathering and trampling, gnawing and scavenging, root etching, and even heat exposure (accidental or intended). Moreover, the body or its parts can be subject to transport by wind, water, animals, or simply to gravity, when it is exposed somewhere uphill. As a consequence, most fossils will be removed from their original context and the place of recovery will differ from the place of individual death. In the case of hominids, this situation changed after humans started to intentionally bury the dead. Finally, the bodily remains become buried, and diagenesis sets in. Diagenetic processes involve all interactions of the dead body or its parts with the surrounding sediment, which either end up in the complete disintegration and dissolution of the body or in its lithification. “Diagenesis” is a term derived from the Earth Sciences and refers to the biological, chemical, and physical changes that take place in sediments after their deposition and that end up in lithification.

In this narrow sense, diagenesis also applies to fossils. Today, this term is frequently used in its broader sense and is synonymous with “decomposition,” and refers also to subfossil and prehistoric bones and to finds of forensic relevance. The state of preservation of a dead tissue will give clues to burial conditions, burial practices, and sometimes even to antemortem conditions such as cause of death or existing diseases. Antemortem injury or infection, or any pathology which causes tissue breakdown during life is likely to accelerate decomposition. Even after lithification, fossils may show up on the surface due to erosion, where they are again subject to weathering and trampling.

Taken together, the taphonomic processes lead to a stepwise loss of information about the formerly living being, or at least to a decline of the integrity of the information through the many ways that the evidence can be altered (Reitz and Wing 1999). However, the statement by Shipman (1981a, p 3) “in short, through death most evidence of the interesting information about animals—what they look like, what they eat, how they move, where they live, and so on—is lost” now requires, more than 20 years later, some revision: methodological progress, particularly in terms of archeometry, permits the evaluation of several of the above-mentioned parameters, including genetic relationships in cases of preserved DNA (this book, Chapters 9, 17, and 20). But it becomes also clear that any research on the molecular or crystal level necessitates deeper insights into taphonomic, especially diagenetic, processes. Thus the early statement by Weigelt (1927) that taphonomy and disintegration will give clues to the paleoenvironment in which the fossils were buried still holds today, and taphonomy has now become an integral area of paleoenvironmental research.

Nonetheless, taphonomy has largely changed from a paleontological subdiscipline into a scientific field of its own, where research mostly focuses on either biostratinomy or diagenesis (Domínguez-Rodrigo et al. 2003). Since the 1980s,
several books covering this topic provide proof of the development of this field of research (Behrensmeyer and Hill 1980; Binford 1981; Brain 1981; Shipman 1981a; Allison and Briggs 1991; Donovan 1991; Lyman 1994; Martin 1999). Taphonomic analyses are also of great importance for archeozoology, since animal bone deposits are usually the remainders of kill or butchering sites, or village or home-based refuse. The animal bones are characterized by several kinds of manipulation before being deposited, and the individual animals have been transferred from the living community into the archeological deposit (Reitz and Wing 1999).

With regard to the aims and scopes of this handbook, this chapter concentrates on the taphonomic and diagenetic processes acting on the vertebrate skeleton, since the mineralized parts of a body have a higher probability of turning into a fossil than soft tissues. However, depending on the mode of preservation, the latter may also leave their traces.

### 7.3 Modes of preservation

Several modes of preservation exist, which are explained in the following sections.

#### 7.3.1 Mineralization

This involves organic matter and necessitates exposure to water. Organic matter is gradually infiltrated and replaced by minerals derived from the watery environment, and mineralization results in the preservation of the shape of the organic component, sometimes down to individual cells (Nicholson 2001).

#### 7.3.2 Carbonization

This results from heat and/or pressure, which removes all volatile organic components and leaves nothing but a carbon layer. This way, coal is generated from dead plant tissue.

#### 7.3.3 Permineralization or petrification

This is the result of the infiltration of natural hollow spaces of plant tissue, such as the xylem of woods, by infiltration and precipitation of exogenous minerals. If this exogenous material is made out of silica, wood will petrify.
7.3.4 Recrystallization

This is a process during which original mineral crystals change in size and shape, mediated by temperature and pressure. The chemical composition, however, is not changed, as in the case of the change of aragonite (CaCO₃, rhombic crystals) into its more stable form, calcite (CaCO₃, trigonal crystals). After death, mineralized tissues may be subjects of a spontaneous rearrangement of the crystalline matrix (Henderson 1987). Since the crystals change in both size and shape, morphological details will not be preserved, and even destruction may occur.

7.3.5 Heteroionic substitution or replacement

It is a mode of preservation, whereby the mineralized components of a body are substituted molecule-by-molecule, or ion-by-ion, with exogenous material. This process also takes a long time because the original mineral has to first dissolve before it can be replaced by exogenous precipitates, like silica, which are less soluble. Morphological details will be preserved, contrary to recrystallization; hence substitution may lead to a “perfect” fossilized skeletal part.

7.3.6 Molds and casts (authigenic preservation)

These are special fossil types wherein a body part gets trapped into sediments which harden, while the original bodily tissue (mostly bone or shell) dissolves. Thus, a void is produced within the sediment, and its negative relief can preserve the impression of the former body part left on the still soft sediment. Hardening sediment, producing a cast (with a positive relief), may again infiltrate this void. It is also possible that sediment infiltrates a shell or a hollow space of a skeleton, e.g., the brain case, and hardens while the shell or bone dissolves. The resulting mold will also have a positive relief as in the case of the brain cast of the Taung child. Under special circumstances, bodies may be preserved unaltered, like insects trapped into amber. A special category however concerns.

7.3.7 Trace fossils

These can be defined as “modifications resulting from the activity of living organisms on and in substrates.” Trace fossils include gnawing and butchering
marks on bones (feeding traces, see later), but mainly concern visible tracks of a no longer visible body (or its parts), like fossil rodent burrows, sleeping pits of hibernating cave bears, or the so-called “bear polish,” a gloss on stones resulting from the polishing action of bears fur in narrow cave passages (for review, see Gautier 1993). The most famous such fossil in the paleoanthropological record is the preservation of the australopithecine footprints from Laetoli. Such impressions will be preserved when they are buried quickly. Rapid burial is less uncommon than sometimes assumed, since not only volcanic ash eruptions, but also storms, floods, or landslides may cause sedimentation within hours.

In this sense, fossilized feces called coprolites (from the Greek copros = dung and lithos = stone) can also be included into this highly variable group of trace fossils because coprolites are the result of the metabolic transformation of living organisms (edible plants and animals) into feces that subsequently fossilized. Coprolites from many animal species including dinosaurs have been recovered and they are mostly preserved by mineralization after rapid burial (Hollocher et al. 2001, Chin et al. 2003). However, very old coprolites are not easily identified in terms of their origin, and this holds especially for the fossilized feces of omnivores. Coprolites suggested to be of hominid origin were removed from the Australopithecus africanus site at the Olduvai Gorge, from the Homo erectus site at Terra Amata, and from a Homo neanderthalensis site at Lazaret. In none of these cases could their hominid origin be confirmed by conventional coprolite identification methods (cf. Trevor-Deutsch and Bryant 1978).

### 7.3.8 Bog bodies

These are not fossils in the narrow sense since they are not lithified. However, the special environmental condition of a sphagnum bog (acidic pH, low temperature, lack of O$_2$) leads to preservation of soft tissues by tanning. Preservation by lithification is thus not a common fate of a dead body and will occur only under special circumstances. As a result, not all places in the world have the same likelihood of bearing fossils somewhere under the surface. For instance, the acidic soil conditions of densely forested areas do not permit fossilization at all. Karstic caves, on the other hand, are characterized by low temperature fluctuations and slightly alkaline conditions, which constitute a nearly ideal environment for fossilization (Weiner et al. 1993). Normally however, complete decomposition is the normal fate of a dead body, and the speed of the processes involved depends on the local burial environment. The tissues will normally suffer from stepwise destruction, and some destruction usually also precedes preservation.
Abiotic and biotic factors may both be responsible for the alterations of dead bodies in the course of necrology, biostratinomy, and diagenesis. These processes are also referred to as “first-order changes” because they are independent from archeological activities and are opposed to “second-order changes,” which are under control of the archeologist, such as choice of site (where to excavate), excavation technique, etc. (Reitz and Wing 1999).

Preburial modification of body parts is brought about by scavenging, gnawing, digestion, butchering, weathering, trampling, transportation, and heat exposure. Postburial modifications are the result of the burial environment, which is characterized by the availability of water and oxygen, temperature, pH, abundance of microorganisms, faunal and floral elements, and ion concentrations. Many taphonomic changes may resemble pathologies or even intended manipulations and give rise to killing or cannibalistic scenarios—another facet of taphonomic research, which is responsible for distinguishing a biological signal from a taphonomic alteration.

The amount of paleobiological information that can be deciphered from a fossil is a function of all three steps of its taphonomic history. For instance, the body of individuals that died of predation have a smaller chance of preservation because the mineralized body parts will undergo at least some partial destruction by chewing and stomach acid etching. In ideal, rare cases, individuals are caught in natural traps like the famous La Brea tar pits in California. Such sites can contain fairly large amounts of accumulated fossil bones (Shipman 1981a). Catastrophic death due to flash floods, volcanic eruption, etc. has to be distinguished from attritional death by natural causes, since the age structure of the resulting bone assemblage will differ drastically. While catastrophic death leaves a sort of snapshot of a formerly existing animal community, attritional death will lead to an overrepresentation of the more vulnerable very old, very young, and diseased individuals.

The majority of alterations to vertebrate skeletons later recovered in fossilized form are due to biostratinomical processes. Thus, the recovery of a nearly complete fossilized skeleton or even of several parts of the same skeleton (e.g., the case of Australopithecus afarensis A.L. 288-1 “Lucy,” or the “Nariokotome Boy,” H. erectus KNM-WT 15 000) is exceptional. Usually, single skeletal elements or their parts are recovered out of their former anatomical context, with teeth constituting the majority of fossil finds. This is the result of the natural disarticulation sequence of a vertebrate body, which is altered and enhanced by predators or scavengers, and which depends on the structure and composition of an individual bone. The over- or underrepresentation of specific skeletal
elements at the australopithecine site of the Makapansgat Limeworks has already been stated by Dart (1957), and was confirmed at the Sterkfontein, Swartkrans, and Kromdraai sites, giving rise to both experimental and field work (Brain 1967). Still, a major concern of taphonomic studies is the differentiation between natural and cultural origins of bone accumulations, and experimentation, the results of which are necessarily interpreted by uniformitarian assumptions, is usually straightforward despite its inherent limitations (for review, cf. Nicholson 2001, Denys 2002).

The mobility of joints and the presence of more resistant tissues like tendons control the natural disarticulation sequence. First, the skull and limb bones are disconnected, followed by the ribs. Next, the limb bones and the lower jaw become disarticulated, only when the final stage is reached, when weathering will mostly have already set in, does the vertebral column become scattered (Toots 1965). Once skeletonized and disconnected, the resistance of individual bones toward further destruction is a function of bone density (Lyman 1984), at this stage, the process of destruction appears to be rather constant and less a function of the exact taphonomic context (Stojanowski et al. 2002). Best preserved are skulls, followed by the shafts of limb bones and their denser epiphyseal parts, while the sternum, clavicle, and scapula are usually the least well preserved. Teeth have a very good chance for preservation because of their high density, small surface/volume ratio, and overall small size, which favors rapid embedding.

7.4.1 Preburial modification processes of body parts

7.4.1.1 Weathering and trampling

When talking about weathering, most people intuitively think of bone exposed to climatic conditions while it is still on the surface and not yet buried. Behrensmeyer in her pioneering fieldwork defined six stages of bone weathering: from stage 0 (no cracking or flaking) to 6 (bones mechanically fall apart into pieces), but made it clear that weathering may destroy a bone in situ “either on surface or within the soil” (Behrensmeyer 1978). While it is a common occurrence that bones buried within the soil may still “weather,” although at a much slower rate, the main interest of paleontologists, anthropologists, and forensic scientists is to define the time a skeletal element has spent on the surface, or how much time has elapsed until a bone assemblage was formed (Potts 1986). In their critical evaluation of Behrensmeyer’s weathering stages and the conclusions drawn from them, Lyman and Fox (1997) confirm the main factors governing bone weathering (size and density, species-specific texture, stability of the
environment—three factors only, but enough to render the process quite com-
plex), but clearly point out that one is still “a long way from safely inferring time
of bone assemblage formation, let alone hominid behavior, on the basis of bone
weathering data.”

In the course of the weathering sequence a bone undergoes while still on
the surface, longitudinal fractures appear first, followed by fractures running
transversely to the shaft of a long bone. Fragments which become abraded, may
resemble bone artifacts like needles and awls. Miller (1975) pinpoints that there is
an “ever-present danger of misinterpreting bone that had been altered by natural
causes as artifacts.” Needle-like bone splinters even with holes, strongly resem-
bling needles at first glance, may be the product of stomach acid etching after
bone ingestion by scavengers (Sutcliffe 1970).

Experimental work revealed that trampling of a bone can produce scratches
by sediment grains that strongly resemble cut marks. Real cut marks, in contrast,
may be completely erased by trampling. This is another example of pseudoarti-
facts which have been mimicked by natural phenomena, and which can lead to
serious misinterpretations of a fossil. It has therefore been emphasized that a
correct identification of cut marks is greatly supported by their location on a bone
(site of ligament or muscle attachments), and by their microscopic features
(Behrensmeyer et al. 1986).

7.4.1.2 Transportation

Although whole bodies can be transported away from the original place of
death by such agents as floods, the likelihood of transport is much higher for
disarticulated body parts. These are taken away by predators and scavengers
or are relocated by wind currents or gravity. However, “water is perhaps the
most ubiquitous taphonomic agent transporting, modifying, and accumulating
remains” (Haglund and Sorg 2002). The authors emphasize that many fossil
hominid finds, such as the Nariokotome Boy or the australopithecine “First
Family,” have been recovered from sand or gravel sediments deposited by moving
water, which are highly suitable for rapid embedding and burial. Since water
environments differ in current, temperature, depth, salinity, and other features,
disarticulation sequences and decomposition processes of a dead body are diffi-
cult to predict. Experimental taphonomic work with human bones in an artificial
flume, which constitutes an ideal and not a natural water transportation simul-
ation, have revealed different transport groups of bones according to their size and
shape (Boaz and Behrensmeyer 1976). This does not only lead to a removal of
skeletal parts from their original place but also to a sorting of the bones in terms
of their morphological properties. Certainly, water transport causes additional damage to bones, especially by abrasion in sediment-rich water.

### 7.4.1.3 Scavenging and gnawing

It is the scavenger’s business to clear dead bodies from a site. Spontaneously, one would think of middle-sized mammals, like hyenas, but also small rodents, toads, and some land snails are also efficient scavengers. Many more species gnaw bones, mostly to meet their calcium and phosphorus demands. These include nonhuman primates, canines, weasels, raccoons, hyenas, cats, artiodactyls, squirrels, beavers, mice, rats, porcupines, rabbits, hares, varanids, and some desert tortoises (cf. Reitz and Wing 1999). Osteophagia is especially common in artiodactyls (cattle, red deer, reindeer, muntjac, camels, and giraffes) as a symptom of phosphorus deficiency in regions with low phosphorus content of parent rocks. Since the bones are chewed with the molars, such gnawed bones reveal a pattern resembling forks and prongs, which may again be mistaken for human artifacts (Sutcliffe 1973). Rodents, which have to wear their ever-growing incisors, leave very characteristic gnaw marks, preferentially on the edges of the bone, in the form of parallel grooves. Bite marks left by carnivores do not occur as parallel grooves but as irregular grooves and pits (Figure 7.2). Four types of tooth marks, namely, punctures, pits, scoring, and furrows, have been identified by Binford (1981). The trabecular parts of the bones may be completely consumed, and the thicker compact shaft will exhibit splinters and sharp, pointed ends (Haynes 1980; Eickhoff and Herrmann 1985). According to Shipman (1981a), carnivore tooth marks have a V- to U-shaped cross section. Bones that have been digested are eroded and pitted by stomach acids. They are usually of very small size and will mostly be recovered only by sieving feces or regurgitants.

Fossil bone assemblages raise the question whether the bones have been accumulated by scavengers or by hunters (natural versus cultural accumulations), especially in the case of early hominids, which are assumed to have practiced a mixed hunting–scavenging subsistence strategy (Shipman 1986). Extensive fieldwork revealed a fairly general consumption sequence of medium to large sized carcasses (Blumenschine 1986): consumers proceed in two cycles from the back to the front of the carcass, consuming, first, the easily accessible, high-yield nutrients of the flesh and viscera, and, second, the more difficult-to-access bone marrow and brain (Figure 7.3). Blumenschine (1986), therefore, postulated a relationship between body part representation and an inverse consumption sequence. Bone assemblages due to hunting consist of more or less complete sets of body parts including those with large, easily accessible, meaty portions,
whereas bone accumulations as a result of scavenging should consist of the less valuable parts of a carcass. A reevaluation of the hominid fossils recovered from Sterkfontein Member 4 (Pickering et al. 2004) in fact revealed that the paucity of postcranial remains, and the carnivore tooth marks on the fossils, strongly supports the "carnivore-collecting hypothesis" as the underlying mechanism for the accumulation of fossils at this site (cf. Brain 1981). Forensic studies

Figure 7.2
Bite marks on the epiphyseal end of a tibia from a sheep/goat. Punctures are produced by canine teeth. Bones from juvenile animals like this one are rather soft and easy to chew. Note the additional cut mark at the right lower end of this fragment. The bone thus shows features of both butchering and gnawing. Photo: S. Bischler

Figure 7.3
General consumption sequence of carcasses according to Blumenschine (1986). Deviations from this general pattern are due to carcass size, age of the dead individual, and consumer type.

First cycle:
- Tissues outside bone (flesh, viscera)

Hindquarter: posterior of rib cage

Forequarter: postcranial incl. rib cage

Head

Second cycle:
- Tissues inside bone (marrow, brain)
also confirmed that canid disarticulation of human bodies follows this rather consistent sequence but may cover highly variable time spans. Total disarticulation was achieved between 5 and 52 months and characteristically, the cranium could always be recovered in contrast to other skeletal elements (Haglund 1997).

### 7.4.1.4 Butchering

Butchering marks are cut marks produced by the tools utilized in the process of dismembering or skinning a carcass (Figure 7.4a and b). While their cross section may be very similar to carnivore bite marks, they frequently exhibit fine, parallel striations (Shipman 1981a, b; Eickhoff and Herrmann 1985) and have specific locations on the skeleton. According to Noe-Nygaard (1989), butchering marks reflect the tool type; hence, cut marks can be distinguished from scrape marks, hack marks, saw marks, and blows. Blows, especially, will fracture the bone and

![Figure 7.4](a) Cutmark on a pig calcaneus. Photo: S. Bischler, (b) Damage to a pig tibia due to butchering that permitted access to the bone marrow. Photo: S. Bischler
serve getting access to the bone marrow. According to experiments carried out by Walker and Long (1977), cross sections through a cut mark permit the identification of the tool, and the shape and depth of cut marks varies with the applied force. Other experiments by Bromage and Boyde (1984) open up the possibility of telling the directionality of cut marks on bone, which would give clues to the handedness of their producers. However, the microstructural properties of the marks are highly dependent on the bone properties (type of bone, bone density, etc.).

### 7.4.1.5 Heat exposure

Excluding cremation as a special burial custom, and all cannibalistic scenarios, food processing is the most probable event when a bone is exposed to elevated temperatures for a prolonged time, and this will hold for animal bones in the majority of cases after hominids became capable of controlling fire. As long as a bone has been in direct contact with the heat source, burning and incineration will leave characteristic traces. Even after very short exposure, when charring is still absent, the denatured collagen will produce both longitudinal cracks (similar to those produced by weathering) and also transverse cracks in the bone, which are the result of recoiling of broken collagen fibers that had been previously under tension (Shipman 1981b). As long as the bone is protected by surrounding flesh during baking, roasting, or boiling, heat exposure remains undetectable macro- and microstructurally. Heating experiments in which whole bovine bones were boiled for several hours revealed that protein loss is accompanied by increased crystallinity and porosity, effects that might cause accelerated diagenetic deterioration (Roberts et al. 2002). If this hypothesis holds true, another bias in terms of the composition of bone assemblages will have to be expected. However, diagenetic processes (see in the following section) may have precisely the same effects.

### 7.4.2 Postburial modification processes of body parts

#### 7.4.2.1 Diagenesis

Diagenetic processes are controlled by the various features of the burial environment, that is the availability of oxygen and water, temperature and its fluctuations, soil pH, the presence of soil flora, soil fauna and especially soil microorganisms, and the availability of free ions (Figure 7.5). Biodegradation can thus be differentiated from chemical degradation, but both pathways are usually intertwined.
Moisture, temperature, and pH are responsible for the abundance of soil microorganisms, which get easy access to a bone by invasion through the natural hollow spaces. Microorganisms consume organic matter to meet their energy demands; therefore, the biological breakdown of a bone’s organic constituents is also far from random (Balzer et al. 1997). Primary metabolic products of this organic matter metabolism are CO₂ and H₂O. In addition, many microorganisms are capable of secreting acid metabolites, which will subsequently dissolve the biological apatite, leaving characteristic destructive foci (Figure 7.6). Many bacteria, like those of the genus Streptomyces, are capable of selectively cleaving collagen by their collagenases. Consequently, collagen is lost from the bone, leaving a higher porosity of the specimen. Together with the focal destruction, the surface area of the specimen is enlarged (Grupe 2001). Even in the absence of microorganisms, soil pH may lead to a slow loss of collagen from a bone by chemical hydrolysis (Figure 7.7). Mineral infilling of the diagenetic porosities is considered a key mechanism leading to bone fossilization (Collins et al. 2002).

In general, it is plausible that microbial invasion of a buried bone plays a key role in bone decomposition, since prolonged microbial attack will prevent fossilization and accelerate breakdown. Since microbial activity is temperature and moisture dependent, warm and wet climates will not be suitable for bone preservation. Exposure to excess humidity on the other hand will reduce the necessary...
Figure 7.6
Microradiograph of medieval human compact bone (Espenfeld, Thuringia), showing intensive focal destruction after microbial invasion

Figure 7.7
Subfossil human bone (Göbekli Tepe, Anatolia, 9500–8500 cal BC). While the microstructure is well preserved with the exception of several microfissures, this specimen did not show any birefringence under polarized light. No collagen was retrievable from this specimen. Photo: S. Dummler
oxygen supply, and microbial activity is suppressed. Microbiological diagenesis is also reduced in cold climates (Grupe 2001).

The enlarged surface area is an important factor in decomposition rates (Gill-King 1997), since the exchange of matter (minerals and ions) between bone and the surrounding soil is facilitated. Exposed bone mineral may be subject to heteroionic substitution, supporting fossilization. Several elements are prone to be taken up by the bone, especially F, U, and rare earth elements (REE), whereby the latter serve as indicators for considerable diagenesis (Metzger et al. 2004). More frequent however, is the chemical dissolution of the apatite by hydrolysis, depending on the soil pH. This will result in free ions, mainly calcium and phosphate ions, which will either be leached from the bone by soil moisture and precipitation or be recrystallized as soon as a critical ion density is reached (White and Hannus 1983). This recrystallization and/or the preferential loss of smaller crystallites in the course of mineral dissolution leads to an increase of the bone’s crystallinity, a most fundamental aspect of bone diagenesis (Person et al. 1995; Nielsen-Marsh and Hedges 2000). Newly formed crystals are generally more soluble than the original bone mineral, and disintegration proceeds. This holds for instance for the transformation of the original bone mineral, a calcium-deficient hydroxyapatite [Ca$_5$(PO$_4$)$_3$(OH)] into brushite [CaHPO$_4$·2H$_2$O] as described by Herrmann and Newesely (1982). However, recrystallization may also lead to more stable mineral configurations and is therefore a means of fossilization.

7.5 Conclusions

As time progresses, the conditions of a burial environment may change considerably. Therefore, any excavated bone specimen exhibits a certain status quo of its diagenetic history, which is hard to unravel in detail. Considering the gross diagenetic processes, it is obvious that certain environments are more conducive to preservation (anoxia, dryness, cold temperatures), while others enhance decomposition. It is a common occurrence that rather young archeological bones may be much less well preserved than older ones. First, the simplistic explanation holds that a well preserved or even fossil specimen must have been buried in a preserving environment. Second, diagenesis should also be considered a self-limiting process in many cases, especially when the burial environments are favorable for mineral infillings, recrystallization, and ionic replacement (Figure 7.5). Therefore, the majority of recovered specimens will have undergone diagenesis to a certain degree. Since progress in methodological innovations today permits the retrieval of even trace amounts of organic molecules like collagen,
noncollagenous proteins, lipids, and DNA, all of which potentially yield very useful, diverse, and otherwise nonachievable information of the once living organism, the understanding of diagenetic processes today mostly focuses on the means and speed of the deterioration of biomolecules. Many studies, therefore, attempt to define easily obtainable predictors for molecular preservation (Götherström et al. 2002).

Acknowledgments

My special thanks are to the editors of this handbook for inviting me to contribute to this volume. Susanne Bischler, MA and Sara Dummler, Dipl Biol kindly provided some of the figures, and George McGlynn, MA edited this manuscript.

References

Behrensmeyer AK (1978) Taphonomic and ecological information from bone weathering. Paleobiology 4: 150–162


Pickering TR, Clarke RJ, Moggi-Cecchi J (2004) Role of carnivores in the accumulation of the Sterkfontein Member 4 hominid assemblage: A taphonomic reassessment of the complete
Weigelt J (1927) Rezente Wirbeltierleichen und ihre paläobiologische Bedeutung. Max Weg, Leipzig
Abstract

This chapter looks at the potential explanatory power of archeology as a paleoanthropological discipline. Through the study of artifacts, archeology provides insight into tool use as a major part of human behavior. Five facets of archeological knowledge can be drawn from the material remains of object-bound activities: typological, technological, functional, contextual, and cognitive information, from which other behavioral aspects, like subsistence, settlement, social organization, and their cognitive background, can be derived. Yet, archeological analyses face significant constraints. A brief outline of Paleolithic artifact history shows how this part of the behavior of *Homo* developed, which is accessible by archeological remains. Following the question “What could they do?” the cognitive basis of that behavior is tracked down in its phylogenetic-biological, ontogenetic-individual, and cultural-historical dimensions. The study of the evolution of the human mind is still at its beginning in archeology. Six different approaches are given as examples.

8.1 Introduction

Archeology is the anthropological field that studies the material cultures of former human populations. The most ancient of these populations are the subject of Paleolithic archeology, which deals with the material remains of early hunter-gatherers, from the first stone artifacts of 2.5 million years ago to artifacts documenting the initial steps toward sedentary life and agriculture at the end of the last Ice Age. Paleolithic archeology is often perceived as studying typical and rather invariant Oldowan, Acheulean, or Middle Paleolithic stone artifact assemblages that sometimes accompany human fossils. Moreover, because archeological analyses of these associated finds has remained fairly dissociated from evolutionary discussions in physical anthropology, the assumption has formed that human design, production, and use of tools stagnated during at least the first 2 million years (Myr) of their existence. In fact, however, Paleolithic artifacts are considerably more informative than this: they provide clues to early human behavior and
attest to hominins’ increasing capabilities and efforts to get beyond the physical limits of their bodies by using tools. Although tool use and manufacture do occur occasionally in the animal kingdom—ranging from the use of unmodified objects in one specific context that some species of invertebrates, insects, birds, and mammals have developed to the tools adaptive to different problem settings that primates, especially great apes, and some bird species, namely crows, tend to use—as will be described later, tools emerge as an eminently human production indeed. For modern humans, who have created an artificial world, tools are their second nature. How we got from our occasionally tool-using primate ancestors to a species that is helpless without tools can be traced only by following the development of artifact assemblages. Therefore, this chapter will focus on the potentials and limitations of interpreting artifacts and tools found in archeological contexts to illuminate this special human character.

8.1.1 Definitions

As a prelude to the study of tools and tool use, three terms must be defined: “archeology,” “artifact,” and “tool.” Archeology, in a broad sense, can be defined as the search for human activity zones, the recovery and documentation of these zones, and the analyses of remains (artifacts) and their interpretation. The main purpose of the discipline is to study artifacts made and used by prehistoric populations in order to reconstruct human material culture and activity patterns and thereby elucidate the development of humans as cultural beings (Figure 8.1). To this end, the raw data from excavations undergo analysis and sometimes additional experimental studies on the manufacture and use of artifacts and on the processes through which sites have developed over time. Insights into the origin and formation of artifact assemblages may also be gained by examining the object behavior of living human groups under ethnoarcheology (Binford 1978) or by comparatively analyzing the material remains of recent primate activity (Mercader et al. 2002). In sum, artifacts constitute the main material basis of archeological studies. Animal and plant remains from human activity areas are also analyzed, and dealt with as artifacts if they were used by human beings, to provide information on environmental parameters that influenced human behavior. Finally, human skeletons can be subject to archeological research insofar as they either were manipulated—and therefore can be handled as a sort of artifacts—or show physical features induced by certain activities.

Artifacts and tools are overlapping but not synonymous categories. While tools and their functional use are the main focus of studies of animal behavior, archeology concentrates on the artificial aspect of manipulating objects in human
context. This point of view is summed up by Hahn (1993), who differentiates among natural objects, items in human context, and things which show signs of human use. Following his definition, the category “artifact” includes all material objects manipulated by humans, from a stone moved by an individual to clearly human-made pits, hearths, and stone structures. More narrowly, Hahn defines artifacts as items of stone, wood, or other materials that show at least some indicative use-wear, whether the objects are separate from or fixed to the environment. The term “tool” bears a technological meaning in archeology as a subcategory of artifacts: they are freely movable objects and were commonly modified in several operational steps, conditions which are not necessary to be accepted as an artifact. Thus an artifact blank can become a tool intentionally by modification—as in the case of stone flakes reworked into scrapers, denticulates, and burins, for example—or unintentionally by use for some functional purpose, as evidenced by use-wear and retouches on an unmodified blade.

In contrast to the archeological viewpoint, primatologist Beck (1980) defines tools not technologically by the object itself but by the functional use to which an object may be put. Tool use, he writes, is “the external employment of an unattached environmental object to alter more efficiently the form, position, or condition of another object, another organism or the user itself when the user holds or carries the tool during or just prior to use and is responsible for the proper and effective orientation of the tool” (Beck 1980 p 10). Thus, Beck
excludes clearly artificial structures, like nests, pits, or other fixed artifacts, which are quite common in animal species and are often highly complex, such as the elaborate nests of bower birds, classifying these as the material outcome of a different, nontool behavior. Ergonomist Christopher Baber (2003 p. 8) identifies a tool as “a physical object that is manipulated by users in such a manner as to both affect change in some aspect of the environment and also to represent an extension of the users themselves. The manipulation is directed toward a specific goal or purpose, and the associated activity requires a degree of control and coordination.” Accordingly, in this functional sense, a simple form to cut out cookies is a tool but even the most sophisticated cookies themselves are not: those are mere artifacts of tool use.

The ethological definition of tool use differs from the archeological one in two ways: tool use is not restricted to human behavior and objects fall into the category of tools by their use, irrespective of the technical aspects of use-wear and modification that characterize archeologically defined tools. Thus Beck, dealing mainly with directly observed behavior, is but marginally concerned with the material evidence of use on the artifacts; and for zoologists generally, who focus on behavior and are only in passing documenting the material inventory of populations or species, the technological details are of minor interest. As a result, a major problem in Lower Paleolithic archeology, that of identifying artifacts and used objects made from durable material, is excluded from the domain of ethologists studying modern animal tool behavior. For archeologists, however, who consider the use and manufacture of objects as a characteristic human means to increase the physically limited abilities of the body in solving problems, the functional separation of tool use sensu stricto from the formation of attached or unattached artificial structures is irrelevant. Both can be subsumed under the category of artifacts to get a more complete picture of the material culture, the subject central to archeology. What sort of information artifacts do yield through archeological analyses will be discussed next.

8.1.2 Facets of archeological knowledge

Through the use of diverse methodological approaches, five major facets of knowledge can be drawn from archeological artifacts. These facets can also be combined to generate pictures of different aspects of prehistoric life like settlement and subsistence behavior, social and religious organization, technical progress, and the spread of innovations.
Typological facet: Artifacts tell us what forms and styles of objects typify a group. Artifact types are defined by features of form or style as perceived by modern analysts, and changes in types through time, as seen through stratigraphic analysis, seriation, or both, form the basis of relative chronology. The definition of cultural groups and their geographical distribution is founded on the relative spatial separation of artifact types.

Technological facet: Artifacts tell us the materials and operational steps used to make tools. By analyzing the different operational steps in tool production, their order, and the variation in that production sequence, we can gain insight into the basic crafts of past societies. Further, a survey of the implements necessary to make certain artifacts gives evidence on the complexity of human object-behavior. In addition to technological studies, the management of raw material is examined to see, for example, if raw material was taken by chance, if certain raw materials were preferred, or if special raw materials were used for particular artifacts. The effort taken to procure the raw material for artifacts and the degree of extensive or intensive exploitation are further relevant to developing hypotheses on how former people differentiated and evaluated technological qualities. A combination of typological and technological aspects of knowledge may refine an artifact typology and can give information about group- or period-specific employment of crafts.

Functional facet: Artifacts reveal not only the technological aspects of how they were manufactured and got their form but also how they were used. Through experimentation, use-wear analyses can indicate the activities and ways in which artifacts were employed. Additionally, functional examination can give evidence about behaviors whose resulting artifacts may be only rarely preserved such as the manipulation of organic raw material (Soffer 2004). Further, experiments with reconstructions, as in the case of the wooden spears from Schöningen/Lower Saxony (Rieder 2003), can test hypotheses on qualities of tools. The examination of specific functions related to special artifact types reveals formal categories already set up at the time of the use of the tools. In all, a combination of functional and technological analyses can disclose the life history of artifacts with sequences of repeated modification and use as well as hypothesize which functional necessities of a tool influenced its manufacture.

Contextual facet: Artifacts tell us, from their location in a site, where they were made and used, and suggest how sites and life within them were organized. Based on detailed three-dimensional documentation of where artifacts, tools, and waste were found, combined with refittings of production sequences, loci of activity can be classified as dump zones, ateliers,
food processing areas, etc. The spatial and chronological relations between the identified activity zones and between contemporaneous sites allow conclusions on the organization of prehistoric life.

- **Cognitive facet**: Artifacts tell us what kinds of decisions either deliberately or unconsciously underlie their conceptualization, manufacture, and use. Thus comparative analysis of data from the four previous facets can reveal preferences in how tools were designed, produced, and used and where on a site they were made or used, pointing up less favored or rejected alternatives. Information on a population’s artifact array, combined with environmental constraints, give hints about the range of behavioral choice and aspects of decision-making in typology, technology, use of artifacts, and site organization. In this way, the spectrum of the decisions taken by a group exposes its knowledge about and comprehension of the world, its cognitive capabilities and the processes of cultural diffusion underlying artifact types, production and use; in rare cases, as at the French Magdalenian sites of Etiolles (Pigeot 1990) and Pincevent (Ploux 1989), even individual behavior and apprenticeship can be perceived.

In these five aspects of archeological knowledge derived from artifacts, it becomes obvious that the archeological discipline is a combination of science and humanities not only in its questions and collected data but also in its approach to knowledge. While the technical, functional, and contextual aspects are scientifically oriented, yield somewhat reproducible facts, and are open to experimental falsification (not in the hard scientific sense but in the sense that, for e.g., wear patterns on a tool can be repeatedly examined, and hypotheses on their origin tested), the typological and, even more so, the cognitive facet adopt a humanities-oriented hermeneutic approach. This attitude of knowledge looks not so much for identification of facts and rules but for understanding of background patterns. Of course, this differentiation between scientific–empirical and humanistic–hermeneutic approaches in archeology can reveal only tendencies: almost all archeological works are a mixture of scientific observations and resulting statements with hermeneutic interpretation. The gradual preference of the one or the other depends very much on theoretical standpoints of the analyst.

### 8.1.3 Limits to archeological analyses

Despite the discipline’s remarkable insights into otherwise undisclosed aspects of everyday life, archeology is able to give only an incomplete and fragmented picture of prehistoric activities and behavior. Several preconditions have to be
met so that archeological analysis may gain access to at least some sectors of hominin living.

First, to become archeologically evident today, behavior in the past must have caused a detectable change in the natural environment, producing artifacts or leaving other material traces behind. This prerequisite severely diminishes archeological studies to a small but nonetheless important portion of the full range of human behavior.

Second, the artifacts must have been embedded in sediment, and the sedimentation should have proceeded quickly, but not aggressively, so that the activity context of the artifacts is sealed carefully. This is a key point, for large parts of the information record can be lost due to taphonomic processes in slow or incomplete sedimentation and to secondary alterations of the find-bearing strata like erosion and bio- or cryoturbation.

Third, the artifacts must have been preserved over time. A wide range of artifacts made from organic material, for example, is likely to have weathered to unrecognizable shapes or to have decomposed altogether. Physical and chemical destruction of different substances depends very much on climatic and soil conditions, with cave sediments in temperate zones being most favorable for many raw materials and the acidic red soils of tropical regions among the worst areas for conservation.

Fourth, the artifacts must be discovered. Further, they have to be unearthed with documentation of their context to reveal the maximum information they still bear. The quantity of archeological evidence in an area or a period is a direct function of the density of research activities and the exposure of relevant geological strata by natural means or through quarrying or construction work.

Finally, the traces of behavior kept in the artifacts must be accessible to interpretation. This can depend on actual research paradigms, the possibilities and availability of analytic methods, as well as individual or social limitations affecting the analyst.

In Paleolithic archeology, another problem arises from the difficulties in ascribing artifacts to one of several contemporaneous human species established by fossil remains in the same region. For example, although the first stone artifacts are widely accepted as the products of early Homo habilis or Homo rudolfensis, there are also robust australopithecine candidates for authorship because of their hand morphology, which, in principle, would have enabled them to knap stones (Susman 1991), and it is unclear who made the cut-marks on bones associated with Australopithecus garhi fossils at Bouri/Ethiopia (de Heinzelin et al. 1999). These uncertain attributions hamper the linkage of behavioral data and information derived from physical–biological context and thus obstruct the deciphering of the evolutionary implications of cultural remains.
8.1.4 Archeology as a paleoanthropological subject

While clues to the behavior of deceased species and populations are limited, functional morphology may give insights into general aspects of living by analyzing dentitions adapted to certain diets, or limbs allowing special ways of locomotion, or other features of the genetically-based layout of the body.

From the physique, behavioral possibilities can be ascertained. A generalized hand, with opposable thumb and the capability of power and precision grip, is able to do what a hoof or fin cannot. Fossil skeletons of Homo and other hominins show bipedal primates developing lower limbs specialized for long walks and free hands perfect for object manipulation. But the question of whether robust australopithecines in fact used their in principle capable hands (see earlier) to produce Oldowan artifacts must remain open, as must the actual use made of the enlarging brains that accompanied the evolution of the genus Homo. Only a few features observed on skeletons can be attributed to concrete activities; in most cases, rather unspecific characteristics, like robustness of the bones, are hard to interpret (Bridges 1995), and these limitations are compounded by the sparseness of the skeletal database and lack of systematic examination of the modern reference. Trauma patterns of Neanderthals apparently paralleling those of modern rodeo riders (Berger and Trinkaus 1995) are a chance result that remains singular.

More productive are correlations between basic physical capabilities, including the cumulative markers of effort, stress, and nutrition that can be deduced from skeletons, and artifacts, which yield detailed information about actual behavioral episodes. This explanatory power of artifacts may be demonstrated with regard to the issue of subsistence, as the prominent behavioral aspect that can be derived from Paleolithic archeological remains. For years, the question has been discussed of whether some of the human ancestral groups were hunting or actually scavenging (Blumenschine et al. 1994). While traces from carcass-processing give evidence of the range of the prey according to species and age distribution, as well as show the use of the different body parts, finds like the spears from Schöningen (Thieme 1997) indicate sophisticated hunting activities among H. heidelbergensis. Similarly, clues to the composition of the diet can be gathered not only from the zoological and rare botanical remains from archeological sites but also from markers detected on artifacts. Thus digging for termites, well known for chimpanzees, may also be likely for Australopithecus robustus based on the use-wear analysis on bone tools (Blackwell and d’Errico 2001); and starch grain analyses of stone tools demonstrate that the selection and processing of barley and wheat was underway in Southwest Asia at least 12,000 years before these grains were domesticated (Piperno et al. 2004). A third
example is the control of fire which reaches back at least 790,000 years at Gesher Benot Ya’akov/Israel, from which certain artifacts show burn marks that can be explained only as exposure to controlled, not natural, fire (Goren-Inbar et al. 2004).

Another important issue in Paleolithic archeology that is accessible through artifact analysis is settlement behavior. Artifacts from a range of settlement sites, from short-term camps with nonspecific structures to long-term dwellings at favorable locations, like Upper Paleolithic Dolní Věstonice, Gönnnersdorf, or Pincevent, yield information on the organization of everyday life and on people’s mobility, group size, differentiation, and separation of activities.

It is clear from these artifact remains that places reserved for special activities developed quite early in human evolution, underscoring that resource management is a key factor in human behavior. There were repeated purposeful visits of early Homo to the 1.6-Myr-old MNK Chert Factory Site in Olduvai Gorge Bed II, a tool manufacturing site with more than 30,000 documented artifacts made from raw material brought in from about a kilometer away and distributed over an area at least a kilometer from the site (Stiles 1991). At the Middle Pleistocene “horse butchery site” Q2 GTP 17 from Boxgrove, H. heidelbergensis brought six to seven flint nodules from the cliff some hundreds of meters away and flaked them into handaxes for the immediate purpose of cutting up the carcass of a horse (Roberts and Parfitt 1999). In analyzing the raw material and artifact transport to and from such sites, the development of larger and more complex settlement systems and territorial organization in human evolution can be brought to light (Féblot-Augustins 1999).

As a paleoanthropological subject, archeology can complete the results of other anthropological disciplines in several ways. In analyzing artifacts, archeology gathers information from short-term behavioral episodes which originate from singular or repeated events and can be fused to a more general picture of the behavioral aspects of chronologically, spatially, or biologically distinctive groups. Furthermore, comparing these results diachronically opens a cultural-historical dimension, thus making the evolutionary perspective on human behavior, and its underlying cognitive development, accessible in its full range. This will be described in more detail in a later section.

### 8.2 What they did: A brief outline of Paleolithic artifact history

As an overview on historical knowledge derived from artifacts, and basis for a succeeding discussion of human cognitive evolution, this section covers the major
behavioral developments which can be extracted from archeological remains; short sketches only, they are organized in roughly chronological order.

The earliest artifacts to which hominin authorship is assigned are stone tools reported from Gona, Ethiopia (Semaw et al. 1997) dated to 2.5 Myr BP. Contrary to earlier assumptions that the initial phases of the Early Stone Age (Lower Paleolithic; the Oldowan and the succeeding Developed Oldowan A technocomplexes) could be characterized by a core tool technology with distinctive tool types as intentional end products, experiments have shown that the Oldowan core tools vary with the raw material used (Toth 1985). However, beside simple technologies with coarse raw material from which few flakes were detached, there are examples of higher-quality raw material processed in longer operational sequences (Kimura 1999) with up to 30 flakes removed, both unidirectionally and multidirectionally, from both natural and prepared platforms (Roche et al. 1999). Flakes, few of which show intentional retouch, served mainly as cutting devices, while cores which were typologically classified in categories, such as choppers and chopping tools, were used as heavy-duty tools. In the first million years of human stone artifact production, raw material transport did not exceed some 5–10 km. No manufacture of bone tools has been proven so far, yet in rare cases the use of bone has been documented (Brain and Shipman 1993). Early sites with Lower Paleolithic artifacts are located in East and South Africa, but by around 1.8–1.7 Myr, the first migrants to West Asia (Dmanisi) (Gabunia et al. 2000) and presumably to East Asia were also been equipped with similar technology. Throughout the Oldowan period, observed site organization is widely unspecific with focus on processing carcasses and/or stone tool production (Stiles 1991). Simple settlement structures have been postulated at Olduvai Gorge, Tanzania, and Melka Kunture, Ethiopia, for example, but these claims are mostly rejected today due to possible natural explanations.

With the Acheulian, a new technological concept of bifacial tools with specific intentional forms, like hand axes and cleavers, was introduced around 1.5 Myr (e.g., Konso-Gardula, Ethiopia) (Asfaw et al. 1992). This technology eventually spread from the Early Acheulian sites in Africa to later sites in Asia and Europe. Although the Acheulian bifacial concept dominates the perception of that period, these artifacts are in fact copresent with flake tools, though these are often, possibly falsely, classified as spatially and/or chronologically distinct industries like the Clactonian and Tayacian in Europe or the Developed Oldowan B in Africa (White 2000). Movius (1949) has observed that East and Southeast Asia seem to have been completely excluded from the spread of the bifacial technology concept, but recent finds from the Bose valley of Southern China (Hou et al. 2000) dating to around 0.7 Myr cast doubt on this hypothesis. Probable evidence of the control of fire as a typical human artifact is increasing
with the 0.79–Myr-old finds from Gesher Benot Ya’akov, Israel (Goren-Inbar et al. 2004), although some instances of fire use, such as at Zhoukoudian /China from the end of the Lower Paleolithic period, have been questioned (Weiner et al. 1998).

From the very end of the Lower Pleistocene on, human occupation has also been proven for Europe, e.g., at Atapuerca TD6, Spain (Carbonell et al. 1995), with human fossils strengthening the weak artifact evidence. Although the material remains of Homo behavior become more numerous in the subcontinent from 0.5 Myr on, it is not likely that the whole of Europe was populated continuously. The observed artifact spectrum widens in the Middle Pleistocene with wooden spears and other tools, e.g., from Clacton (Oakley et al. 1977), Schöningen (Thieme 1997), possibly Bilzingsleben (Mania and Mania 1998), and Gesher Benot Ya’akov (Belitzky et al. 1991). A few flaked bone tools have been documented in that period but no bone or ivory projectile points (Villa and d’Errico 2001). Evidence of nonfunctional artifacts, such as at Berekhat Ram (d’Errico and Nowell 2000), Hunsgi (Paddayya 1977), and Bilzingsleben (Mania and Mania 1988), is very scarce. Postulated settlement structures at Bilzingsleben (Mania 1983) and the Grotte du Lazaret (de Lumley 1969) are questioned.

In the final phase of the Lower Paleolithic, and continuing into the African Middle Stone Age and the European Middle Paleolithic, flakes were increasingly retouched to improve or create the working edges of tools like scrapers, denticulates, notched pieces, and special forms of knives. The rising focus on flake tools can also be traced in Levallois technology, a new concept in stone knapping that first appeared between 0.3 and 0.2 Myr. The idea of characteristically prepared Levallois cores, which allow predetermined flake forms, did not replace the older concepts but was used in addition to them. In the European Middle Paleolithic, distances of raw material or tool transport increase up to 100 km or, in some cases, more (Féblot-Augustins 1999). Artifact assemblages of the Upper Pleistocene after the last interglacial also reveal increasing complexity and diversification. Microwear patterns and use of hafting materials, like a piece of birch tar at Königsaue, Germany (Grünberg et al. 1999) or bitumen at sites in the Levant (Boëda et al. 1996), indicate the manufacture of composite tools, while stone tool assemblages in both Africa and Europe show increasing regional, chronological, and perhaps functional variation. Also on both continents, in Late Middle Paleolithic/MSA context, there appear bone tools like intensively-used awls (d’Errico et al. 2003) and bone points (Henshilwood et al. 2001), as well as the first instances of personal ornaments (Henshilwood et al. 2004) and the application of coloring material and decorative elements (Henshilwood et al. 2002). From around 0.1 Myr, there is evidence of intentional interment of human bodies; whether these qualify as burials and indicate symbolic behavior is still
under discussion. Finally, evidence of hearth structures, for example at Kebara (Meignen et al. 2001), mark a more habitual use of fire.

The Early Upper Paleolithic, starting at around 40,000 years and continuing until around 30,000 years ago, is characterized by intensified use of bone, antler, and ivory as raw material. In some regions of the world, a burst of personal ornaments and artistic representations is documented, e.g., rock art as in the French Grotte Chauvet (Valladas et al. 2001) and Australia or mobile art (Conard 2003) and musical instruments (Conard et al. 2004) in caves of the Swabian Alb. Although blades had been elements of the artifact inventories since the Lower Paleolithic, a new technological concept of blade production widely introduced in that period allowed an extended exploitation of the stone cores by optimizing the number of rather uniform blanks. In addition, Upper Paleolithic assemblages show increasing diversification in specific tool types for distinct purposes. Settlement of the European Arctic (Pavlov et al. 2004) and, most likely, of Japan took place in this time period; Australia had become inhabited slightly earlier (O’Connell and Allen 2004).

In the Middle Upper Paleolithic, from about 30,000 to roughly 20,000 years ago, the first unquestioned evidence of true burials with symbolic meaning, e.g., at Sungir (Pettitt and Bader 2000) and Dolní Věstonice (Klima 1988), are documented. Tool types became more and more standardized and regionally diversified. The period is known for an important extension of the find spectrum: evidence of fiber processing (Adovasio et al. 1998), systematic gathering of grass seeds (Piperno et al. 2004), production of unfired and fired clay sculptures (Klima 1991; Einwoegerer 2000), and, for the first time, construction of large and complex settlement structures mainly in the mammoth steppe, for example at Dolní Věstonice and Pavlov (Klima 1991). While depictions of animal–human hybrids and elements of the human body, like vulva signs and hand negatives, are known from the Early Upper Paleolithic, the Middle Upper Paleolithic introduced representations of entire human beings, which were extremely rare before and now became part of the parietal and mobile art spectrum in frequent Venus figurines, a few male images and some portraits. In this period, human occupation was extended to arctic Siberia (Pitulko et al. 2004).

The Late Upper Paleolithic, after the last glacial maximum, saw the introduction of complex tools that work only in combination with other artifacts like spear throwers with spears or needles with thread. A new category of artifacts, domesticated animals and plants, started to develop in human cultural contexts. Humans accompanied by the first domesticated animal, the dog (Sablin and Khlopachev 2002), migrated from Siberia to the Americas (Leonard et al. 2002), crossing the Beringian landbridge.
In conclusion, this brief outline, although only a superficial picture, demonstrates a gradual developmental process of progress, diversification, and intensification in archeological sites from 2.5 Myr ago until the end of the Pleistocene. It is important to note, however, that all observations of this period are biased by the decreasing chances, over time, of artifact conservation. Furthermore, summary views of large time periods and spaces may appear clear-cut but, if examined in detail, become blurred.

### 8.3 What could they do? Recovering the basis

Advanced tool behavior, as observed in great apes and the artifact record of *Homo*, is founded to only a minor extent on instincts—genetically based processes automatically released by key stimuli—and is instead derived from cognition. Thus, like all cognitively controlled behaviors, it can be stopped arbitrarily, altered by learning, and improved by experience.

The basis of cognitive behavior is the individual cognitive space (or mind), which can expand along three dimensions (Figure 8.2): the phylogenetic, the ontogenetic, and the historical-cultural. The phylogenetic dimension of an individual’s or a species’ mind expresses the genetic potential derived from evolutionary processes. Phylogenetic development in human cognitive evolution can be assumed, since great apes demonstrate marked differences in cognitive behavior, regardless of whether they grew up in an intact natural group of their own species or received individual cognitive support in a modern cultural-historical environment. The cause of this development in humans may be related to
physical features like the increase of relative brain size (McHenry and Coffing 2000) and possible changes in brain anatomy (Eccles 1989). Genetic studies suggest other or additional possible agents, like increased gene activity in the brain (Enard et al. 2002a) or the mutation of FOXP2, a gene involved in linguistic articulation (Enard et al. 2002b); their specific roles, however, remain unclear. The exploitation of the phylogenetic potential cannot ever be assumed to reach completion because it depends on the two other factors, the ontogenetic-individual and cultural-historical dimensions; therefore, even if two individuals or populations have the same phylogenetic cognitive potential, their cognitive performance as seen, for example, in artifacts may be completely different according to their cultural background and individual experiences.

The ontogenetic-individual dimension incorporates the cognitive elements originating in individual actions and experiences, be they accidental or intentional. This dimension is limited by the biological potential to think and act, given by genetic characters derived from phylogeny and is influenced by opportunities to interact with the environment. Behavioral innovations, which do not directly descend from genetic mutation, originate in this individual dimension. For the expression of the ontogenetic-individual dimension, the frame is set by the phylogenetic and the cultural-historical dimension, although individual performances may lie outside the population’s mean range which is normally extrapolated from archeological data.

The cultural-historical dimension, finally, expresses the culturally fostered factor of the mind, the cognitive dimension that is most prominent in and, in its full range, exclusive to humans. This dimension is opened by the spread of innovation in behavior, not by genetic transmission, between genetically related and unrelated individuals, within and between generations. On a low level of this cultural dimension, individuals adopt a certain behavior, whose outcomes they observe in others, by emulating it until they are personally content with the result. In a real cultural setting, however, with teaching and learning between individuals and shared attention on a problem, children do not have to find solutions on their own for problems that arise but can rely on culturally stored solutions, traditions, invented by individuals in past decades, centuries, or millennia. This historically grown solution set makes up a part of the individual’s environment that can be acted on and used as a basis for further innovation—the so-called Ratchet effect (Tomasello 1999). The cultural-historical dimension of the mind does not expand constantly; instead this dimension and with it the spread of innovations are strongly influenced by interdependent social parameters like intra- and intergroup communication, population density, social structures, the position of innovators in their groups, and general group-specific attitudes about
learning, innovation, and progress (Rogers 1995). Factors, such as communication, that may hamper or foster the increase of the cultural dimension have their origins partially in the phylogenetic dimension, probably with language faculty or the capability to understand others as intentional actors (Tomasello 1999).

In sum, the development of the human mind and the cognitive aspects materialized in artifacts should be viewed not as an exclusively phylogenetic process with some saltational breaks in its linear progress but as an exponential expansion in the three dimensions of cognitive space, the exploitation of which is incomplete and variable. Thus, early and rare or singular expressions of a capacity like symbolic behavior—in the Late Acheulian figurine of Berekhat Ram, for instance—can be seen as proof of the phylogenetic and ontogenetic ability of the actors at that time. As a communal trait within a specific cultural-historical context, these individual expressions may only be accepted and reproduced thousands of years later.

Cognition is the main basis of what prehistoric groups did, and this is partially expressed (Section Archeology as a paleoanthropological subject) in the artifacts that have been preserved to our times. Thus artifacts are a means for detecting the cognitive background behind their creation. This is not easy. Indications of prehistoric people’s cognitive potential—what they could think and do—have to be separated out from behavior compelled by the restrictions of the natural and social environment. Archeology can help to delineate the cognitive space of prehistoric groups and trace the development of the cognitive dimensions, especially the cultural-historical factor. Yet loss of evidence within the archeological record must also be factored in; it must be kept in mind that absence of evidence cannot be equated with evidence of absence and in not only material but also cognitive terms (Speth 2004). What we can detect in the archeological record is only a group’s minimum cognitive potential, which has been manifested in artifacts. Cognitive faculties that are apparently unexpressed in material remains because a group did not represent these faculties, or because an archeological analyst failed to recognize them, might have been present; yet researchers can only then state the lack of indication.

8.3.1 How to extract cognitive capacity and evolution from archeological remains?

The evolution of the human mind is an old question in paleolithic archeology (Verworn 1915). The flourishing of cognitive science (Varela 1990), beginning in the 1950s, did not touch the discipline’s potential, however: for a long time archeology and cognitive science widely ignored each other. While Leroi-Gourhan
(1964) sketched a singular developmental picture of the paleontology of human thought, most other archeological enquiries into cognition were predominantly raised to redefine human uniqueness, after widely recognized studies of animal tool behavior, documenting several animal species’ use of stones as tools (Hall and Schaller 1964; van Lawick-Goodall and van Lawick-Goodall 1966; Boesch and Boesch 1983) demonstrated that tool use could no longer be considered an exclusively human marker. The use of a tool to produce a tool, however, seems to represent, a specific human capability. This secondary tool use (Kitahara-Frisch 1993)—e.g., the use of a stone to alter another stone so it could be used to solve a specific problem—is a universal human trait for which there is only sparse and questionable evidence in chimpanzees (Sugiyama 1985, Takemoto et al. 2005). Today, regarding typically human cognitive performance and its origin, the focus in archeology has narrowed to so-called nonfunctional, symbolic artifacts related to “modern” human behavior as art, religious actions, and indirectly language, combined with planning and foresight (Noble and Davidson 1996; Mellars 1996; Klein 2000; Coolidge and Wynn 2001).

In fact, the wide disregard in Paleolithic archeology of the evolution of cognition is primarily based on the prevailing theoretical paradigms. New or Processual Archeology, while open to evolutionary approaches, considers cognition, with few exceptions, inaccessible to scientific analyses and testing. Postprocessual Archeology, although amenable to examining human thought in archeology, has rejected evolutionary approaches as simplistic, biologistic, disregarding of the cultural and individual bases of human behavior, and conducive to artifact assemblages which can hardly be generalized. Outside these mainstreams of late-twentieth-century archeological theory, there have been some attempts, described later, to approach cognitive evolution archeologically under the theoretical concept of Cognitive Archeology, whose approaches derive mainly from psychology.

### 8.3.2 Paleontology of human thought

One archeological attempt to illuminate the evolution of cognition appears in “La geste et la parole,” a pioneering work by French paleontologist and prehistorian André Leroi-Gourhan (1964), who saw the development of the modern human mind paleontologically as a process in which the physical evolution of the brain released new capacities for thinking. The early development of bipedal locomotion, for instance, not only freed the hands for technical actions but, with the shift of the foramen magnum, also enabled an increase in size of the occipital lobes. Subsequent simple technical solutions, like primitive stone tools, allowed
the reduction of the masticatory apparatus, which made further change in skull architecture possible; specifically, the parietal and especially the frontal lobes could gain their modern anatomical forms and structures, and the mouth, with decreased nutritional functions, could be used for vocal communication. Brain volume is of secondary significance to Leroi-Gourhan; besides the anatomical features of an upright posture, a short face and hands not engaged in locomotion, the main marker of humankind has been the manufacture and use of tools.

Leroi-Gourhan posits that human cognitive evolution was dominated by increases in technical intelligence and vocal capabilities, capacities that developed slowly until H. erectus, and then became more sophisticated, although still mainly technical, in Neanderthals. Only with the unfolding of the prefrontal cortex, bringing capacities for foresight, consciousness, the control of affect, and the ability to discern and thus to reflect on behavior, could late Neanderthals and, later, anatomically modern humans develop nonfunctional actions, symbolic capabilities, and creative consciousness. Thus, development was not bound solely to biological evolution, with memory encoded predominantly in the genes and with individuals inventing technical operations on their own. Rather, the new cognitive capacities fostered the ethnic group as a social memory pool with true language as a cultural storage facility, allowing the development of operational sequences in super-individual processes: in teaching and sharing problems and solution attempts with other members of the groups. The result was the explosion of artifact varieties and symbols in the Upper Paleolithic. In sum, Leroi-Gourhan’s focus lies on the physical basis of cognitive evolution with secondary support by archeological evidence. His model concentrates on the expansion of the phylogenetic dimension of cognitive space, with the cultural-historical factor becoming important only at the end of the Middle Paleolithic.

8.3.3 Early Homo capabilities in the light of Piaget’s theory

A second and completely different perspective on human cognitive evolution has been taken by Thomas Wynn (1979, 1981, 1985), who applied Jean Piaget’s model of ontogenetic developmental stages in children’s object behavior to Lower Paleolithic artifacts and thereby to human phylogeny. The first developmental stage, young infants’ sensorimotor intelligence, characterized by pure activity-based intelligence without inner representations of the actions, can also be observed in primate tool behavior; in this stage, activity cannot be reasoned out in advance. The second stage, preoperational intelligence, is marked by inner representations of single consecutive tasks, so that anticipation of the results of an
action is limited to the change in only one variable at time; planning of an action is therefore restricted to trial and error. Wynn (1981) has identified this stage in chimpanzee tool manufacture, as well as in the simple technology of Oldowan core tool assemblages (for the problems of Oldowan interpretation see (8.2)).

The next stage, the concrete-operational phase, allows coordination of changes in several variables. It is now possible to anticipate the result of an operational sequence or to construct the operational sequence for reaching a desired result, so complex planning can proceed, and errors can be envisioned and corrected before they are executed. This concrete-operational stage, which according to Piaget should be fully developed by age 11–13, is identified by Wynn (1979) in the bifacial technology of Acheulian hand axes. Diverging from Piaget (1970), who differentiated between concrete-operational intelligence in children and the formal-operational intelligence that succeeds it in modern adults, Wynn merges these two stages and concludes that human phylogenetic development of cognition—as it can be seen from technological perspective—reached modern competence around 300,000 years ago, at the latest. Although Wynn’s interpretation needs reevaluation in light of challenges to Piaget’s theory by modern psychology, his attempt to adapt an object-directed psychological theory to archeological analysis of artifact traits is a positive example for future studies.

8.3.4 Looking for the origin: Toward primate archeology

A third means used to illuminate the phylogenetic dimension of the cognitive basis of early human tool behavior has been to make comparisons with ape performance. In the last years, the knowledge of tool behavior in great apes has markedly broadened through research on chimpanzees Pan troglodytes (McGrew 1992; Whiten et al. 1999) and orangutan Pongo pygmaeus (van Schaick et al. 2003), as well as on the less-observed bonobo Pan paniscus (Hohmann and Fruth 2003). Probable cultural differences among groups have been compiled and correlated with complex object-behavior, and these data have expanded archeologists’ view of the range of culturally influenced behavior and artifact categories not preserved in archeological context. As noted previously, however (Section Definitions), primatological results answer archeological questions about basic object behavior only partially.

To fill this gap, which is especially critical with regard to stone tool technology, experiments have been designed to examine motor and cognitive abilities in stone flake production. While early testing with an orangutan (Wright 1972) had a behavioral focus for understanding the functional aspects of cutting tools, the long-term experiments with bonobo Kanzi (Toth et al. 1993; Schick
et al. 1999) have been planned in cooperation with archeologists and observed from a technological perspective. Kanzi proved able to detach flakes from a core with a hammerstone, although he preferred to produce cutting edges by throwing a stone on a hard surface, a technique he invented himself. After 3 years of knapping experience, Kanzi could still not flake efficiently by exploiting acute angles on the core, a method regularly observed in Early Lower Paleolithic context (Roche et al. 1999).

Archeologists have also analyzed and compared the stone tools at a chimpanzee nut-cracking site with Oldowan assemblages (Mercader et al. 2002). These comparisons suggested that, despite previous doubts (Wynn and McGrew 1989), morphological analysis could clearly discriminate between human tools flaked intentionally and chimpanzee artifacts produced as a chance result of other activities with stones. Primate archeology is still in its beginnings and has not yielded any model of human cognitive evolution; nonetheless, it is another promising approach for taking a detailed look at the specificity of human tool behavior.

8.3.5 From domain-specific intelligence and to a fluent mind

A fourth approach has been developed by Steven Mithen (1996), who, based on psychological models of different cognitive domains in modern humans, has created a hypothetical model of the phylogenetic development of human cognition. Mithen identifies four distinct domains of intuitive intelligence in modern children that, he believes, reach back to Pleistocene times: linguistic, social, technical, and natural history. Only when these types of intelligence are combined can the knowledge specific to one domain be generalized to the other domains and made usable beyond the circumstances in which it arose, enabling creativity to develop. Analyzing the technical and environmental behavior of modern chimpanzees, as well as the assumed behavior of the common ancestors of humans and chimpanzees, and assessing that both originate in a basic general intelligence domain, Mithen detects in these species only one specialized domain, social intelligence, which he believes became fully developed in early Homo at around 2 Myr. In his opinion, the technical and natural history domains arose sometime up to 1.8 Myr and allowed the Homo genus’ expansion out of Africa and spread to large parts of the Old World. For the following 1 Myr, mental organization did not show major changes, paralleling a trend of stagnation in relative brain growth. Mithen expects that linguistic intelligence came into being between 0.5 and 0.1 Myr but was initially bound to social intelligence; he assumes that technical and environmental issues could not be talked about at that time.
because these issues were still encapsulated in their own domains. A cultural explosion took place between 60,000 and 30,000 BP, when, according to Mithen, the knowledge stored in specific intelligence domains became accessible to the other domains. This cognitive fluidity is evidenced in archeological artifacts like bone and antler tools, blade technology, grave goods, the expansion of occupation to Australia, personal ornaments, and art—all of which require a combination of knowledge from social, linguistic, technical, and natural history domains. This model, while provocative, consists of open generalizations that fail to account, for example, for how the Late Acheulean wooden spears from Clacton or Schönningen could be manufactured before the development of a cognitive synthesis. Thus, Mithen’s model of distinct intelligence domains becoming fluent on the eve of modern behavior can be summed up as adapting hypotheses from evolutionary psychology to human cognitive evolution by matching them in a speculative way to selected parts of the archeological record.

8.3.6 Language and symbolic behavior as milestones of cognition

In a fifth prospective on the cognitive basis of early human tool behavior, Noble and Davidson (1996) erect a phylogenetic model of human cognitive development on the premise that, unlike chimpanzees, modern humans can share their awareness of objects and events, to which they can give socially constructed meaning with the help of language. Their research starts with observations of aimed throwing as the hypothetical germ of pointing, a primeval mode of sharing awareness intentionally. Aiming or pointing, they posit, could have developed into iconic, representational gestures that, in their repetition, could have left chance traces on some material. Once recognized and imitated as representatives of the gesture, these signs might have become symbols; and once the principle of representing symbols was understood, the faculty of language might have arisen as a result of the shared meanings. Language, then, allowed ways of thinking which had been impossible without it: it permitted the description of perceptions, reflections, memories, plans for the future, and awareness of oneself as part of the world and as a perceptive individual. To support their speculative model of the development of shared assignment of meaning through language, Noble and Davidson look for evidence of symbolic behavior and planning in archeological remains. Up to and including the Middle Paleolithic, they find that no specific artifact forms were produced in a goal-directed process and that the observed progress is only technological, not conceptual. For the authors, however, Upper
Paleolithic art and the colonization of Australia by boat are the most prominent among the first markers of modern cognitive aspects which derive from language, constituting true manifestations of symbols, and planning. This model, again, constitutes an attempt to parallelize ideas derived from psychology with some selected artifacts and events in prehistory. Yet this has been done in a rather imaginative way basing more on common knowledge than on strict theoretical discussion and without an empirical analysis of the distinctive traits of artifacts chosen from the archeological data set to support or disconfirm the hypothesis.

8.3.7 Increasing the problem-solution-distance

A sixth approach to human cognitive evolution on the basis of tool behavior focuses on a particular attribute of artifacts: the dissociation of problems and solutions. In his famous experiments, especially with chimpanzees, in the early twentieth century, Köhler (1963) recognized tool behavior as an extension of the process of round about thinking. If tool use is considered to solve a problem, then the immediate desire, to get a fruit for example, has to be set aside for one or several intermediate objectives such as finding or producing an appropriate tool. Thus, at least in the short term, thinking and resulting physical actions must depart from the immediate problem. Separated from each other, the different elements of mental or physical operations may not seem to be useful; if one is unable to conceptualize a tool for termite-fishing, it does not make sense to look for a thin and flexible twig in the bushes 5 m away from the termite mound. However only in the combination of these different operational elements, the final aim can be reached. In essence, thinking shifts from the immediate givens of a problem to abstract conceptualizations of possible solutions and sequences of means, necessitating appropriate tools, for achieving these solutions in the future. The more that tools and their manufacture can be dissociated from immediate subsistence aims, the more problems become soluble. In handling artifacts, humans are able to speculate on the thinking that may have guided the artifacts’ manufacture and original purpose as well as to think about possible improvements in the tools and adaptations to other problem contexts. Thus, tools can not only be used mechanically but also serve as cognitive instruments in metareflections—in past as well as future directions. By producing tools, humans do more than just make the environment increasingly manageable; they also create a material and cognitive world on their own, which expands with the increasing dissociation of subsistence constraints, perceived problems, and available solutions.
An increasing dissociation of problems and solutions, as well as an expansion of the range of problems perceived beyond subsistence, can in fact be observed in the development of human artifact behavior (Haidle 2000, 2004, 2006). In Oldowan context, the transport of raw material over some kilometres and the use of secondary tools (tools produced by employing tools to solve a problem) both indicate an awareness that problems may require a phased solution over time, and that provision must be made in the present for solving problems in the future. Extended operational sequences, which must follow a consecutive although not necessarily strict plan, can be observed in bifacial technology, assuming that the knapping of bifacial tools represents the shaping of an intentional form. Further, a marked extension of production time and use of tertiary tools—that is, tools manufactured by the use of a secondary tool—is present in Lower Paleolithic wooden spears, for example, since a stone tool had to be prepared as an intermediate instrument to carve these means for hunting.

In so-called nonfunctional, symbolic tools, such as art objects and ornaments, the link between problems and solutions becomes exclusively conceptual. Additionally, the problems these tools solve go beyond individual needs to challenges relevant to the social community. With the help of true symbolic tools as communication devices, information can be transferred by the mere existence of these artifacts in a certain context without any further action of their creator. An additional expansion of the problem–solution distance becomes evident in organized trade and specialization on selected activities: solutions are offered for other people’s needs, a behavior which is in the following extended to abstract needs of unknown customers. In short, the range of problems addressed widened slowly over time from direct and personal physical problems to indirect, superindividual, abstract, and cognitive problems, culminating today in products like computers, whose elements, produced independently all over the world, receive their problem-solving power only in combination.

The origin of this phenomenon of an increasing dissociation between subsistence, perceived problems and activities generated to solve them cannot be assigned to one genetically based character alone, but must result from the expansion of the cognitive space in all three dimensions, the phylogenetic-biological, the ontogenetic-individual, and especially the cultural-historical which fosters the cultural storage of solutions. Due to the complex interactions of the three cognitive dimensions with the phenomenon of increasing problem–solution dissociation, the extension of this dissociation accelerates exponentially in its developmental history. This may be at least one reason why only a few, slow changes can be observed in early artifact assemblages, compared to more frequent changes in later artifact history.
8.4 Conclusions

Through the study of artifacts, archeology provides an insight into tool use as a major part of human behavior. Five facets of archeological knowledge can be drawn from the material remains of object-bound activities: typological, technological, functional, contextual, and cognitive information, from which other behavioral aspects, like subsistence, settlement, social organization, and their cognitive background, can be derived. The only available evidence, although fragmented, on the development of human behavior comes through diachronic comparison of artifact assemblages; and thus, questions of how prehistoric populations used their cognitive potential, composed of phylogenetic-biological, ontogenetic-individual, and cultural-historical dimensions, can be approached only through archeology. The study of the evolution of the human mind is still at its beginning, however. Some archeological investigations of human intellectual development set up cognitive traits linked with tool behavior and try to find them in the archeological record, while others take hypothetical psychological models without clear links to tool behavior and try to match them with selected artifacts and events in prehistory. But until the signal strength of archeology, the analysis of tool-related behavior, is integrated into models of human cognitive evolution, these models will remain speculative.

References


Haidle MN (2000) Neanderthals—ignorant relatives or thinking siblings? A discussion of the “cognitive revolution” at around 40,000 B.P. In: Orschiedt J, Weniger G-C (eds) Neanderthals and modern humans – discussing the transition: Central and Eastern...
Europe from 50,000 – 30,000 B.P. Neanderthal-Museum, Mettmann, pp 275–286


Leroi-Gourhan A (1964) La geste et la parole. Albin Michel, Paris


Verworn M (1915) Die Entwicklung des menschlichen Geistes, 3rd edn. Gustav Fischer, Jena
Wright RVA (1972) Imitative learning of a flaked stone technology – the case of an orang-utan. Mankind 8: 296–306
9 Contribution of Stable Light Isotopes to Paleoenvironmental Reconstruction

Julia Lee-Thorp · Matt Sponheimer

Abstract

In this chapter, we focus attention on the potential linkages between hominin evolution and environmental change in Africa, as shown by the stable light isotope evidence. We begin with an overview of the principles and the materials typically available to produce this kind of stable isotope data. Carbon isotope analyses of faunal enamel, soils, and ostrich eggshells provide good evidence for the emergence of C_{4} grasses ca. 5–8 Ma, broadly concordant with hominin bipedalism. Although C_{4} grasses remained a modest yet variable component of the vegetation for millions of years, the associated biomass may have quickly become an important foraging resource for hominins. Rather than a climate-driven “pulse” ca. 2.4–2.8 Ma, evidence from East and South Africa suggests a significant change to more open, grassy ecosystems ca. 1.8 Ma, broadly concordant with the emergence of early Homo. On present evidence, there appears to be only a weak overall aridification trend for the entire period from ca. 4 Ma. Speleothem and other evidence points to deep-rooted cyclicity in floral and rainfall shifts, which suggest that obliquity and precessional orbital cycles were important drivers of environmental variability. In relation to climate and environmental influences on hominin evolution, a revised “Savanna” hypothesis may yet be compatible with the “variability selection” hypothesis.

9.1 Introduction

It is widely accepted that understanding the environments that hominins inhabited provides important information for understanding evolutionary trends. First, and perhaps most obvious, we need this contextual information to tell us about the kinds of opportunities available for the basics of life–food,
water, and shelter, and the biological and perhaps cultural responses used to obtain these things. Second, there are questions about how environmental and climate shifts might have influenced the course of hominin evolution and, similarly, that of other coeval fauna. Here, paleoenvironmental studies must take a broader, more long-term view. Large-scale, global climate shifts are very often invoked as major influences on evolutionary pathways, but in practice demonstrating the links is not at all straightforward (Feibel 1997). This is partly because of the nature of the environmental evidence on land, which is patchy in time and space and often poorly dated, and partly because climate is just one of a number of forces acting on landscapes.

One of the ways to address this problem is to take a broad, continent-based perspective that can “subtract” more local environmental influences. At the same time, we must bear in mind that behavioral (and biological) responses of hominins occur at the local level, which is the most important for understanding the lives of individuals. In other words, scale is important, depending on the question that is being asked. In this chapter, we focus our attention on paleoenvironments of Africa for the obvious reason that much of early hominin evolution occurred there.

A whole range of investigative tools is available for reconstructing paleoenvironments at the site level. Standard methods rely on geological context, faunal abundances, and in some cases floral indicators such as woody materials or pollen. None of these approaches on its own is sufficient because each has its own set of advantages and disadvantages. For instance, methods based on vertebrate faunal-abundance data have to rely on what is deposited, preserved and found in a site so that the fossil bones represent a highly filtered representation of the original vertebrate fauna. Furthermore, in interpreting abundance data as environmental proxies, one has to rely on knowledge of the habitat requirements of various animals, and this is difficult for unfamiliar, extinct taxa. In fact, we very often do not understand enough about the habitat requirements of their modern relatives (Cerling et al. 2003; Sponheimer et al. 2003). Stable isotope approaches can help to provide this complementary evidence.

There are even greater challenges in reconstructing evolving continental climates and landscapes beyond the site level to regional or global contexts. They are partly related to the enormous heterogeneity of continental environments, and partly to the rarity of long, continuous records on land. To this is tied a problem of chronological control. Global climate changes are mainly understood from marine sediment cores that provide an oceanic perspective on climate evolution (Zachos et al. 2001), but it is not always clear how these changes have affected continental African environments. Moreover, it is unclear whether
large-scale climate shifts are synchronous across the continent; between the north and south, and east and west, many recent and ancient climate fluctuations are thought to be opposite (or antiphased) (Nicholson and Entekhabi 1986; Partridge et al. 1997; Tyson et al. 2002). Continuous marine sediment records near continental margins have provided some direct information about African climates. For instance, episodes of windblown dust off North Africa indicate increasing Saharan aridity from ~2.6 to 2.8 Ma (deMenocal 1995), and entrained pollen indicates the pattern of contraction and expansion of West African equatorial forests in the Late Pleistocene (Dupont 1999). These patterns, however, are not necessarily representative of trends for East and southern Africa, where most of the hominin sites are found. Thick sedimentary sequences with diatomites and other organic derivatives from continental lake systems provide good proxies for lake histories and salinity, but they are extremely discontinuous in space and time. For instance, deep, old, but discontinuous lake systems are known from the East African Rift, while none occur in southern Africa. Lake formation is dependent both on basin formation, defined by tectonic activity, as well as on a positive precipitation/evaporation balance, controlled by climate (Trauth et al. 2005). Other land-based clues come from pollen (Bonnefille 1983, 1995) and faunal abundance data considered regionally (Vrba 1985, 1988; Behrensmeyer et al. 1997; Reed 1997; Bobe et al. 2002). Vrba’s “turnover-pulse” hypothesis that postulates a direct link between evolutionary trends in bovids and climate shifts at ~2.4–2.6 Ma (Vrba 1985) was based on bovid abundances, while studies that tested or disputed this idea were based on consideration of radiations among a broader range of fauna (Behrensmeyer et al. 1997; Bobe et al. 2002).

Stable light isotope tools can help to address some of the gaps. These biogeochemical tools have several advantages, because once the principles of isotope fractionation in biogeochemical pathways are understood, they may be broadly applied to past ecosystems to produce quantitative, repeatable results. Stable light isotope applications to paleoenvironmental reconstruction have been extensively reviewed in the literature and the field is broad (for a global review see Koch 1998). Given the breadth of the topic, our intention here is to restrict discussion to some of the key examples that we believe have made, or look set to make, substantive contributions to understand the environments associated with hominin evolution. These contributions lie principally in the context of shifts in floral composition (such as the relative abundance of C4 grasses in the flora, and implications for climate) and climate indicators including temperature, aridity, and seasonality. The nature of longer-term climate evolution can be deduced from these data. We concentrate on examples reliant on materials from hominin or paleontological sites, or from sequences that are closely associated in space and time.
9.2 Isotopic environmental indicators

This section provides a brief overview of the main isotopic environmental indicators and the sample materials that are used to provide indications of prevailing conditions. It is important to understand these principles as they apply to each potential archive, as they are all likely to provide slightly different perspectives, and in order to evaluate the quality of the evidence.

9.2.1 Basis of stable isotope tools

The stable light isotopes of principal interest as environmental proxies are the isotopes of hydrogen, carbon, nitrogen, and oxygen. In all cases, the chemistry of the different isotopes of an element remains the same as chemical properties are controlled by electron configuration. Among these light isotopes, the mass difference owing to one or two extra neutrons is sufficient to cause a small, but predictable, difference in the rates of chemical reactions or physical processes. If the starting materials and the products of a reaction are partitioned in some way, observable isotopic fractionation occurs.

9.2.1.1 Hydrogen (D/H) and oxygen (\(^{18}\)O/\(^{16}\)O) isotope fractionation

Fractionation of hydrogen (D/H) and oxygen (\(^{18}\)O/\(^{16}\)O) isotopes\(^1\) in nature follows similar patterns because their isotope effects are dominated by the processes undergone by water evaporation, condensation, and freezing. The magnitude of fractionation is controlled largely by temperature in both cases, but isotope effects are much higher for hydrogen because of the greater relative mass difference between deuterium (D or \(^2\)H) and hydrogen (\(^1\)H) compared to oxygen (\(^{18}\)O and \(^{16}\)O). Water vapor evaporates largely from low- to midlatitude ocean surfaces, and sea surface temperature (or latitude as a proxy) of the oceanic source influences isotopic ratios of moisture-laden weather systems (Dansgaard 1964). Further influences on the isotopic composition of rainfall occur during subsequent

\(^1\)Isotope ratios are expressed by convention in the \(\delta\) notation, in parts per thousand (‰) relative to a standard: \(\delta X (\%) = (R_{sample}-R_{ref})/R_{ref} \times 1000\), where \(R\) is the isotopic ratio. V-SMOW (Standard Mean Ocean Water) is used as the reference for D/H and \(^{18}\)O/\(^{16}\)O in water, V-PDB (Peedee Belemnite) for \(^{18}\)O/\(^{16}\)O and \(^{13}\)C/\(^{12}\)C in carbonates and organic materials, and atmospheric N\(_2\) (AIR) for \(^{15}\)N/\(^{14}\)N.
transport of these weather systems across continents, related to factors such as the
distance traveled, mountains crossed, and height and temperature of rainclouds
(Dansgaard 1964; Rozanski et al. 1993). A further effect in arid landscapes is due
due to evaporation. Soil- and groundwater isotope values reflect this history, as do
carbonates precipitated from these waters, with the additional influence of local
temperature as fractionation during carbonate precipitation is temperature depen-
dent. This is the basis of the familiar “temperature equation” applied to many
carbonates (e.g., forams, corals, and speleothems) (McCrea 1950). In addition,
isotope effects also occur in plants. A large isotopic enrichment occurs during
daytime evapotranspiration in plants under hot, dry conditions (Yakir 1992).

9.2.1.2 Carbon ($^{13}$C/$^{12}$C) isotope fractionation

Carbon ($^{13}$C/$^{12}$C) isotope fractionation provides fundamental information
about pathways in the terrestrial carbon cycle (see also Sponheimer and Lee-
Thorp Vol. 1, Chapter 17). A large depletion in $^{13}$C occurs during assimilation of
atmospheric CO$_2$ in photosynthesis, but to different degrees depending on the
photosynthetic pathway (Smith and Epstein 1971). C$_3$ plants (trees, shrubs and
forbs, and temperate or shade-adapted grasses) are more depleted in $^{13}$C and have
low $\delta^{13}$C values, while C$_4$ plants (mainly tropical grasses) are less depleted
(\textit{Figure 9.1}). A smaller group of succulent plants follow the Crassulacean acid
metabolism (CAM) pathway, which essentially alternates use of these two path-
ways by night and day, and depending on conditions, so that their isotopic
compositions vary considerably. These plant values are archived in slightly
different ways in a range of sample materials (\textit{Figure 9.1}).

9.2.1.3 Nitrogen ($^{15}$N/$^{14}$N) isotope ratios

Nitrogen ($^{15}$N/$^{14}$N) isotope ratios reveal pathways in the nitrogen cycle, but since
the pathways are complex, the values observed between different ecosystems are
highly variable. Nitrogen enters the terrestrial foodweb via N$_2$-fixing bacteria in
soils or plants to form nitrates or ammonium ions that are utilized by plants. The
net effect of this biological nitrogen fixing and subsequent nitrogen loss is a slight
enrichment in $^{15}$N, but the balance and mode of fixing and loss is strongly
affected by environmental conditions (Heaton 1987). In a global survey, Handley
et al. (1999) found a negative correlation between moisture availability and both
leaf and soil $\delta^{15}$N. Stepwise trophic enrichment of about 3–5‰ occurs in animals
because there is an $\sim$3‰ fractionation between diet and tissues. The trophic
effect, however, can be surpassed by enrichment associated with environmental aridity (Heaton et al. 1986; Sealy et al. 1987; Johnson et al. 1997), effects that are certainly related to raised leaf and soil values in arid areas (Heaton 1987). The effect is very marked in animals living in areas with $<400$-mm rainfall/annum, so $\delta^{15}N$ can be an indicator of aridity.

### 9.2.2 Sample materials

The sample materials most often used as environmental indicators in sites associated with hominin activities consist largely of bones and soils in close chronostratigraphic association with those sites. Fossil ostrich eggshell and cave speleothems represent promising but more rarely explored sources of paleoenvironmental and climate information.

#### 9.2.2.1 Bones and teeth

Bones are abundant in most sites where conditions are conducive to preservation. They conserve information about many of the processes and conditions to which the animal was subjected when it was alive, which can be accessed via their stable light isotope compositions. Bones and dentine consist of about 25% (by weight)
of a fibrous protein, collagen, and about 75% embedded bioapatite mineral. Collagen can be relatively easily purified and analyzed for δ^{13}C and δ^{15}N. However, it tends to decompose over time so that by about 10–20 Ka collagen has all but disappeared from bones and teeth in most African sites (preservation is better in cool Eurasian sites). The mineral is composed of calcium phosphate crystals with many other ions substituted into the structure (LeGeros 1991) that tend to increase their reactivity and solubility. One exception is fluoride, which enhances stability. Substitutions also affect how well bones and teeth are preserved as fossils. Enamel is far more stable than bone apatite (Lee-Thorp and van der Merwe 1987) and consequently most isotopic studies of fossil fauna have relied on enamel as sample material. The component ions of interest are phosphate (PO_4^{3–}) and carbonate (CO_3^{2–}), with the latter occurring in small amounts (3–6%) as a substitution. Therefore, the isotopes available for study in this system are carbon and oxygen isotopes, and the latter may be extracted from either PO_4^{3–} or CO_3^{2–}.

Bioapatite phosphate δ^{18}O has been developed as a paleotemperature tool, based on the rationale that body water δ^{18}O (δ^{18}O_{bw}) in mammals is related to environmental or drinking water (δ^{18}O_{w}), which can in turn be correlated with latitude and temperature effects on rainfall (Longinelli 1984; Luz and Kolodny 1985, 1989). In low- to midlatitudes, temperature effects are far less important than storm-track and amount effects on δ^{18}O of environmental water plus isotope effects on plant waters. We now know that animal behavior, related to drinking patterns and/or thermophysiology, can modulate these environmental signals considerably (Bocherens et al. 1996; Kohn et al. 1996; Sponheimer and Lee-Thorp 1999b, 2001) and that these patterns are conserved in fossil assemblages. Ayliffe and Chivas (1990) found a correlation between δ^{18}O_{PO4} and relative humidity in non-obligate drinking animals—in their case, kangaroos, so that these behaviors might still provide useful environmental data. In order to circumvent the problems of variable responses to climate factors, many authors have relied upon “well-behaved,” obligate drinking species such as equids for extracting paleoclimate proxies (Sánchez-Chillon et al. 1994; Bryant et al. 1996).

Stable carbon isotope (δ^{13}C) analysis of herbivore tooth enamel (or collagen) indicates the relative amounts of C_3 plants and C_4 grasses consumed. The calculations are based on our understanding of typical plant values today and how they might be affected by climate variables, and by the difference between diet and enamel δ^{13}C values (or Δ_{diet-en}). There is a little “play” here, since C_3 plants may be slightly enriched in ^{13}C under hot, arid conditions, and vice versa. C_3 plants in dense forests, where a “canopy effect” with recycling of CO_2 and low light prevails, are typically depleted in ^{13}C (van der Merwe and Medina 1989, 1991). Alterations in the species composition of C_4 grasses, in response to conditions, may shift their average δ^{13}C value slightly (by perhaps 1‰).
In addition, $\Delta_{\text{diet-en}}$ is not precisely established; some authors observe $+12\%$ (Lee-Thorp and van der Merwe 1987), while others have observed up to $+14\%$ (Cerling and Harris 1999). Still, these small differences make no practical difference to distinguishing between C₃ and C₄ consumers since these $\delta^{13}C$ distinctions are large (Figure 9.1). The presence of C₄ grazers in an assemblage allows us to deduce presence of tropical C₄ grasses in a landscape. Although this information alone may be useful, for instance, if a shift between cold- and warm-season rain patterns is suspected (Lee-Thorp and Beaumont 1995), in practice we need to take this process further. Usually we require information about how closed or open vegetation cover was. A C₃/C₄ index has been developed to reflect the relative openness of an environment from $\delta^{13}C$ data in a faunal assemblage (Sponheimer and Lee-Thorp 2003) (see below).

9.2.2.2 Ratite eggshells

The eggshells of ostriches and earlier ratite taxa are durable and often abundant in paleontological and archeological sites. The shells consist of a protein matrix and inorganic calcium carbonate. Eggshells can yield $\delta^{13}C$, $\delta^{15}N$, and $\delta^{18}O$ proxy information reflecting conditions during the short egg-laying season. Ostriches are mixed feeders that prefer tender plants, which can include succulents as well as annual grasses and forbs in their diets; hence $\delta^{13}C$ data can indicate presence of both C₄ grasses and CAM plants (Stern et al. 1994; Ségalen et al. 2006). The $\delta^{15}N$ values are thought to reflect mean annual rainfall (MAP) (Johnson et al. 1997) in the same way as mammals in arid regions with under 400-mm rain/year (Sealy et al. 1987). $\delta^{18}O$ data also reflect aridity but from a slightly different perspective; since ostriches are drought-tolerant animals that do not need to drink free water, $\delta^{18}O$ primarily reflects leafwater enrichment due to evapotranspiration under hot, dry conditions (Lee-Thorp and Talma 2000). Because of the high variability in $\delta^{18}O$ and $\delta^{15}N$ data, large numbers of analyses are required (Lee-Thorp and Talma 2000). To date, only one study has produced a sequence, extending beyond the Holocene based on all three proxies (Johnson et al. 1997).

9.2.2.3 Soils

A history of overlying vegetation may be preserved for long periods in remnant organics and pedogenic carbonates in paleosols (Cerling et al. 1991). The principles are by now familiar: carbon isotopes in both the organic and inorganic components
of soils systematically reflect differences between vegetation following the C\textsubscript{3} and C\textsubscript{4} pathways, allowing estimations of the relative mix of woody and grassy plants on the landscape. $\delta^{13}C$ of paleosol organic matter, where preserved, is a reasonably direct reflection of the mean isotopic composition of the vegetation, with a small enrichment in $^{13}C$ owing to decomposition and associated processes (Cerling et al. 1991). Pedogenic carbonate nodules are formed about 0.5–1 m below the active soil horizon from soil-respired CO\textsubscript{2} (Cerling and Quade 1993). There is a net enrichment in $^{13}C$ of 14–17‰ due to combined effects of diffusion and isotopic fractionation during carbonate precipitation, causing increases of 4.4–5‰, and 9.5–12.5‰, respectively (Cerling et al. 1988, 1991; Cerling and Quade 1993). Pedogenic carbonates have a relatively restricted distribution; for one, nodule formation is associated with semiarid to arid conditions. An important constraint is the identification of “true” pedogenic nodules formed well below the active, identifiable soil horizon, where diffusion enrichment is complete and there is no mixing with atmospheric CO\textsubscript{2} (Cerling et al. 1988; Cerling and Quade 1993).

9.2.2.4 Cave speleothems

Cave flowstones and stalagmites are composed of calcium carbonate formed from CO\textsubscript{2}—rich seepage water dripping into cave systems, degassing, and precipitating as carbonate. Their $\delta^{13}C$ and $\delta^{18}O$ values reflect the proportions of C\textsubscript{3} and C\textsubscript{4} plants in overlying vegetation, and the isotopic composition of the water, in a similar manner to pedogenic carbonates. Speleothems have some decided advantages. For one, a closed cave system is a protected environment, where averaged annual temperatures are maintained year-round, along with high relative humidity. In a closed cave system equilibrium conditions are more likely, meaning that greater confidence can be placed in the isotopic data as sound indicators of environmental conditions. Very importantly, they are incremental structures and can be precisely dated using thorium–uranium disequilibrium or lead isotopes. Ages are more difficult to obtain for older Pliocene- or Miocene-age speleothems, but paleomagnetic and lead–uranium methods are being developed (Hopley 2004). So, where the right material can be located and the ages determined, speleothems can be sampled at small intervals to yield continuous, high-resolution records of a quality unlike any of the other proxies discussed here. Continuous sequences like this are invaluable because they can be compared with other dated continuous records, and they allow us to check trends and the scales of variability.

The following section describes some important applications of isotopic tools to paleoenvironmental and -climate problems.
9.2.3 Emergence of C₄ grasses

One important application has been to document the global expansion of C₄ grass systems between ~8 and 5 Ma (Cerling et al. 1997; Jacobs et al. 1999). Since both biochemical and fossil evidence suggest that our common ancestor with chimps lived at about this time, this important environmental shift might well have had some direct impact on the origins of the hominin lineage. The phenomenon was first detected in pedogenic carbonate nodule isotope studies from East Africa and Pakistan (Cerling et al. 1988; Quade et al. 1989; Cerling and Quade 1993). In East Africa, Cerling et al. (1988) observed the first appearance of C₄ grasses in the Turkana Basin ca. 7 Ma. This pattern of first appearance of significant amounts of C₄ has been mirrored in δ¹³C of fossil fauna in Africa, North and South America, and Pakistan (Cerling et al. 1997), although Fox and Koch (2003) have suggested earlier presence of minor amounts of C₄ biomass in the North American Great Plains, based on pedogenic carbonate evidence. Kingston et al. (1994) and Morgan et al. (1994) have argued that modest proportions of C₄ grasses were present earlier in the Mid-Miocene of East Africa, based respectively on continuous heterogeneity in a pedogenic carbonate sequence from the Tugen Hills and several slightly enriched faunal values from early Tugen Hills sites. The argument hinges largely on the weight of evidence, since the results for the Tugen Hills pedogenic sequence have not been replicated elsewhere in Africa. The values used to demarcate the certain presence of C₄ in fauna also differ as Morgan et al. (1994) used a value of -10‰ as the cut-off point, above which they deduced presence of C₄, whereas Cerling et al. (1997) used a more conservative -8‰.

Most evidence, however, still suggests that C₄ grasses first began their expansion in lower latitudes at ca. 7–8 Ma, expanding over the next few million years to midlatitudes (Cerling et al. 1997). The timing is unclear in southern Africa. Ratite eggshell data document emergence of C₄ and differentiation of the δ¹³C records between the southern and northern Namib about 5 Ma (Ségalen 2003; Ségalen et al. 2006). A large undated, collapsed cone speleothem in the Makapansgat Limeworks in South Africa indicates relatively invariant C₃ vegetation cover from the δ¹³C record, suggesting that it was formed prior to C₄ grasses reaching this region at ~27°S (Hopley 2004). A notable exception to the general picture of C₄ expansion is the 5-Ma-year-old site of Langebaanweg in the southwestern Cape of South Africa (Franz-Odendaal et al. 2002), where faunal δ¹³C data show that a winter-rainfall, Mediterranean-type ecosystem was already in place in this region in the Late Miocene and Early Pliocene. In the northeastern interior, faunal evidence suggests a minor presence of C₄ vegetation by ~4 Ma or earlier from the Rodent Corner at Makapansgat Limeworks (Hopley et al. 2006), while the modest proportion of pure C₄ grazers in the Makapansgat Member 3
faunal assemblage indicates clear presence by $\sim$3 Ma (Sponheimer and Lee-Thorp 1999a).

The expansion of C$_4$ grasslands across large parts of the world must have a global driver but the exact causes have remained curiously elusive. Cerling et al. (1997) and Ehleringer et al. (1997) proposed that plummeting CO$_2$ levels in the Late Miocene, to a level below 500 ppm, favored C$_4$ plants. The hypothesis is based on the known tolerance of C$_4$ plants for lower levels of $p$CO$_2$. But evidence from marine cores suggests that CO$_2$ levels were already low prior to this period (Pagani et al. 1999). This, along with $\delta^{13}$C data from fossil ratite eggshells in Namibia that closely track the marine-derived trends in $p$CO$_2$ for the Miocene (Ségalen et al. 2006), has lead to a reconsideration of the possible drivers. The emergence and spread of C$_4$ may reside in a combination of tectonic and solar insolation forces that rearranged the earth’s global heat budget, both spatially and seasonally. Resolution of the problem likely requires a good deal of more detailed evidence from the transition period, spread across different regions of Africa and elsewhere.

### 9.2.4 Evolution of open African environments

A great deal of interest has been shown in issues related to the decline of large-scale forests and opening up of the African landscape. Some years ago, the “Savanna hypothesis” held that forest shrinkage/savanna expansion in Africa was a primary driver of hominin bipedalism because it was thought that these two trends occurred at about the same time. This hypothesis fell from favor once it became apparent that bipedalism emerged before 5 Ma among relatively wooded habitats (WoldeGabriel et al. 2001). Nevertheless, there have undoubtedly been important and influential changes in the vegetation structures across Africa. Vrba’s turnover-pulse hypothesis relied on the observed radiation of open-country grazing bovids to draw inferences about shifts to more open landscapes in the Pliocene and Pleistocene, which could be linked to hominin evolutionary changes (Vrba 1980, 1985, 1988). This approach is not entirely straightforward because the habitat preferences of extinct bovids, or indeed of any extinct animals, are not always clear. Vrba proposed that a shift to bovid lineages with open-country preferences between about 2.4 and 2.6 Ma was associated with the onset of Northern Hemisphere glaciation observed from marine oxygen isotopes in marine sediment core records (Shackleton et al. 1984). But the nature of the climate trigger is not clear. The idea of a pulse in faunal change has been challenged, although it is acknowledged that faunal changes occurred between 2 and 3 Ma (Behrensmeyer et al. 1997; Bobe et al. 2002).
Further perspectives on the appearance of savanna landscapes have emerged from several isotope studies. Although C₄ grasses appeared in the Late Miocene and persisted with fluctuations through the Pliocene (Cerling et al. 1988; Cerling 1992; Kingston et al. 1994; WoldeGabriel et al. 2000; Wynn, 2004), importantly, they did not become a consistently dominant part of the floral biomass until ∼1.8 Ma (Figure 9.2). A shift in δ¹⁸O also occurs at about 1.8 Ma, suggesting that the source area and vapor transport pathway changed (Figure 9.2) (Cerling et al. 1988). 

Figure 9.2
Isotopic indicators of open, grassy environments from the Koobi Fora Basin and from the South African hominin sites, showing small but variable presence of grasses prior to 1.8 Ma, and a distinctive shift to open habitats after this time. The sequences are as follows: (a) pedogenic carbonate δ¹³C, (b) δ¹⁸O from the Koobi Fora Basin, shown alongside, and (c) calculations for the percentages of browsers and grazers from Makapansgat and Sterkfontein. The arrows provide some guidance about the predicted ranges for pure C₃ and C₄ for (a) and (c), and some pointers to interpretations of the record in (b). Data for (a) and (b) are from Cerling et al. (1988), and they are expressed relative to an age model calculated relative to depth data in the same source. The data for (c) comes from Sponheimer and Lee-Thorp (2003), and Luyt and Lee-Thorp (2003).
et al. 1988). The $\delta^{13}C$ and $\delta^{18}O$ sequences for the Turkana basin shown in Figure 9.2 also illustrate the variability of the results for each period, indicating a rather broad view of vegetation structure. Additional data sets from a wider range of regions and periods have been completed (Kingston et al. 1994; Sikes et al. 1999; WoldeGabriel et al. 2000). These data show a rather complex and variable overall mix of wooded terrain that tends to deemphasize the contributions of C$_4$ grass to the overall mix.

Isotopic studies of fossil faunas, on the other hand, can be designed to investigate the various components of ancient flora. A number of isotopic data sets are now available for Pliocene and Pleistocene faunas that demonstrate presence or absence of C$_4$ grasses in East and South Africa, (Morgan et al. 1994; Cerling et al. 1997; Sponheimer and Lee-Thorp 1999; Lee-Thorp et al. 2000; van der Merwe et al. 2003), and in Chad (Zazzo et al. 2000). In the Chad case, the faunal isotopic data demonstrated significant grassy vegetation from the Late Miocene through the Pliocene (Zazzo et al. 2000). In East Africa, faunal isotope data have in some cases run counter to the perspective afforded by pedogenic studies. At Olduvai, for instance, the isotope data from fauna suggest a greater C$_4$ biomass (van der Merwe et al. 1999) than suggested by the soil carbonates. Another contradiction occurs in the Awash of Ethiopia, where de Heinzelin et al. (1999) suggest a wooded environment at $\approx$2.5 Ma based on presence of certain key fauna (Colobus and Tragelaphus) and soil isotope values, but faunal isotope analyses suggest a significant proportion of C$_4$ grazers (Levin et al. 2004). These discordances suggest that different parts of the ecosystem are being sampled. The advantage of faunal isotope data is that they are abundant in the actual sites associated with hominin activity, and because we can follow first principles to establish abundance of C$_4$ without assumptions about dietary preferences.

No pedogenic carbonate isotope record exists for the regions associated with the South African hominin sites, and the sites are not as old as many in East Africa. The nature of the karstic infill sites present serious challenges for determining the nature of long-term vegetation change. Faunal $\delta^{13}C$ data from all the more important sites analyzed to date (Makapansgat, Sterkfontein, Swartkrans, Kromdraai) have shown that C$_4$ grasses were present from the earliest periods associated with hominins, about 4 Ma or earlier, onward (Lee-Thorp and van der Merwe 1987; Sponheimer et al. 1999; van der Merwe et al. 2003; Hopley et al. 2006). On their own, these data demonstrate presence of C$_4$ but are not very informative about changes in proportions of woody or grassy cover. A “C$_3$/C$_4$ index” has been developed, which is essentially an isotopic expression of the “alcelaphine + antilopine” index developed by Vrba (1980); it is based on the proportions of genera (Sponheimer and Lee-Thorp 2003), or individual specimens (Luyt and Lee-Thorp 2003), falling into one of the grazer, mixed feeder or
browser categories as determined by $\delta^{13}C$. The underlying idea is that browsers will be favored in a closed habitat with many trees, while grazers will be favored in open, grassy landscapes (Sponheimer and Lee-Thorp 2003). When applied to a series of sites to construct a sequential view, the results suggest that the most significant shift toward open, grassy landscapes occurred about 1.6–1.8 Ma (Figure 9.2) rather than at $\sim$2.4–2.6 Ma.

Isotopic data from the Buffalo Cave speleothem, in the Makapansgat Valley, provides more detailed information about vegetation change in this important time period. The speleothem is dated by independent means to between 2 and 1.5 Ma (Hopley 2004). The detailed $\delta^{13}C$ sequence shows cyclical fluctuations in the proportions of C$_4$ grass, with a dominant variability at $\sim$40 Ka, indicating that the major control is the angle of the Earth as it orbits the Sun, and known as the obliquity cycle. Precessional cycles clustered around 20 Ka, due to the Earth’s “wobble”, are also visible. This sequence shows a $\sim$1.7 Ma shift to higher proportions of C$_4$ as before, but importantly, that this was but part of a cyclical pattern. The entire speleothem shows a slight long-term trend toward grassier conditions. Variability in the $\delta^{18}O$ (reflecting rainfall) seems to be most strongly controlled by precessional cycles, in good agreement with those shown for the Tswaing Crater Lake Late Pleistocene sequence (Partridge et al. 1997), and for some Pliocene East African lakes (Trauth et al. 2005). Other stalagmite isotope data from Botswana (Holmgren et al. 1995) and Cold Air Cave in the Makapansgat Valley (Holmgren et al. 2003; Lee-Thorp 2004) hint at considerable floral and rainfall fluctuations but these records are too intermittent or too short to demonstrate precessional cycles. Although both $\delta^{13}C$ (reflecting vegetation) and $\delta^{18}O$ (reflecting rainfall) in the Buffalo cave speleothem are orbitally controlled, the differences in dominant orbital forcing modes between them emphasize the complexities of climate and environment, and the need for multiple sources of evidence.

The shift to grassier conditions in South Africa about 1.6–1.8 Ma is entirely concordant with the shift to open ecosystems in East Africa at the same time. This might be one occasion in which environmental shifts are in phase in East and South Africa.

These changes in floral composition, which can be documented using isotopic tools, are important for evaluating competing hypotheses about links between environment and hominin evolution. What does the floral information tell us? For one, the broad correspondence between the emergence (first appearance) of C$_4$ grasses and hominin bipedalism still holds, in spite of observations that locally, environments may have remained relatively closed. The underlying connection might be dietary rather than locomotor. $\delta^{13}C$ data on hominin diets have, almost without exception, shown an involvement with C$_4$-derived foods.
The implication is that early hominins chose to make use of these new food sources, even if C₄ grass remained a relatively minor component of the ecosystem. We need to test how far back in time this dietary flexibility goes.

There is widespread agreement on the appearance of more open, grassy habitats around 1.6–1.8 Ma. Given the increasing evidence for presence of a fully bipedal hominin equipped with more sophisticated stone tools (including hand-axes) from about this time, it is tempting to draw links between these two occurrences. In this case, we can suggest with some confidence that there is a real environmental “shift,” but the exact nature of linkages to hominin behavioral and evolutionary shifts still require further investigation. Finally, the evidence for cyclical floral and moisture changes from the speleothem data, if upheld, requires reappraisal of several interpretations we have made. These have to do with the issue of time-scale and the nature of the fossil record. Most of the existing data rely on material that can, in principle, reflect small time-windows but the chronologies with which we have to work are quite gross. As a result, what emerges is rather a lot of noise and only the really large-scale environmental trends. The speleothem data also forcibly raise the issues of variability and cyclicity, which are crucial to the “variability selection” hypothesis advanced by Potts (1996, 1998).

### 9.2.5 Aridity indices from oxygen isotopes

Africa never experienced the large temperature swings associated with glacial to interglacial shifts at higher latitudes, and the climate variable of most importance and influence seems to have been rainfall. An overall trend toward aridification has been inferred from the visible trends toward more open, grassier African landscapes over the last ~5–6 Ma, as discussed earlier. However, although aridity and forest diminishment may well be linked in certain respects (for example, maintenance of equatorial forests requires high rainfall), the two are not necessarily tightly connected in savanna habitats. In their report on the isotopic data from pedogenic carbonates from Turkana from ~4 Ma, Cerling et al. (1988) note that the presence of these carbonates through the sequence is indicative that arid conditions were present from the lower sections. Since the δ¹³C data of pedogenic carbonates and fauna in the region shows that the proportions of C₄ grasses remained fairly modest for millions of years after their first appearance, these data suggest strongly that tree density/open landscapes and aridity should be delinked. Studies of modern savanna ecosystems show that lower rainfall often favors more thicket and bushland rather than grass (Scholes and Walker 1993; Owen-Smith 1999). It remains important to establish
independent indicators for aridity. A number of isotopic approaches have been, or are, in the process of development.

As illustrated in Figure 9.2, the Turkana Basin pedogenic carbonate $\delta^{18}O$ sequence suggests that a change in East African rainfall patterning occurred around 1.8 Ma. Since $\delta^{18}O$ values prior to this time were more negative, and those thereafter more positive, the suggestion is that the dominant moisture source changed from a distant (perhaps the Atlantic Ocean) to a closer source, likely the Indian Ocean. The $\delta^{18}O$ values in the latter period are sufficiently enriched to suggest that processes of evaporation also exerted an influence (Figure 9.2). This kind of information, therefore, provides some hints about development of aridity, but little detail.

Isotopic ratios from ostrich or ratite eggshells show great promise as aridity indicators, where the potential is to extract subtle and quantifiable indications of moisture shifts in the past. Johnson et al. (1997) used dual $\delta^{15}N$ and $\delta^{18}O$ data from ostrich eggshell in the Equus Cave hyena accumulation to document moisture fluctuations associated with Late Glacial to Holocene climate shifts. However, no similar studies have been undertaken for earlier periods of interest for hominin evolution, possibly because of the difficulties associated with the organic fraction in older material. Fortunately, data from modern and Late Pleistocene–Holocene archeological sites show that $\delta^{18}O$ alone provides a reasonable aridity index (Lee-Thorp and Talma 2000; Ségalen 2003). $\delta^{18}O$ from the eggshells of earlier ratites has been used to show continuously high, but fluctuating aridity in the Namib from the Miocene to the Present (Ségalen 2003; Ségalen et al. 2006). The main constraint is the chronology, which of necessity is based on biostratigraphy and thus provides only rather crude time intervals. A related tool based on fossil tooth enamel is under development (Schoeninger et al. 2003; Levin et al. 2006), which employs $\delta^{18}O$ differences between animals that use environmental water, and those that rely on plant water and in turn reflect the influences of relative humidity (and aridity). Both of these tools hold potential but have not yet yielded information of direct consequence for addressing questions about climate trends and hominin evolution.

9.3 Conclusions

We have attempted to place some of the long-standing questions about the possible linkages between hominin evolution and environmental changes in an isotopic context and on a sound footing. There are a number of principles and constraints that must be considered for these applications or the results will be of limited value. In doing so, we may have rather overemphasized the problems and
constraints. So, in our concluding remarks, we would like to emphasize some of the solid advances that have been made over the last couple of decades.

A great strength of isotopic approaches to paleoenvironments clearly lies in the area of delineating floral changes, most particularly the presence and proportions of C₄ grasses in the plant biomass. There are a couple of relatively minor disagreements evident from some of the pedogenic versus faunal approaches, but in general, a clear picture emerges. C₄ grasses emerged about 6–8 Ma, probably earlier at the lower latitudes (Cerling et al. 1997), and they remained a visible, but modest and fluctuating, component of the vegetation for several millions of years. Sometime around 1.8 Ma, grasslands took off to become really major components of many African ecosystems, and this change occurred in both East and southern Africa. It is at this point that we probably best see the birth of a modern “world” complete with the vegetation and faunal distributions that we associate with todays “typical” African environments. At the same time, new evidence points to the cyclical nature of environmental change for both flora and rainfall, which is strongly related to orbital forcing cycles of obliquity and precession. These cycles appear to have some considerable time depth. Isotopic evidence for aridity and possible aridification trends is at present thin, but the tools for developing these lines of evidence are now available.

What can these trends, shifts, and cycles tell us about hominin adaptations and evolutionary pathways? Are these issues linked, and if so, how? As noted earlier, the Savanna hypothesis fell from popularity sometime ago because the emergence of bipedalism was pushed back in time beyond estimations for the emergence of savannas, and because immediate environments associated with many early hominin sites suggested closed rather than open conditions. δ¹³C data suggest, however, that the hypothesis might still hold some substance albeit in a rather different sense. For one, the emergence of C₄ grasses and bipedalism is still broadly concordant. The δ¹³C data show that these grasses remained relatively modest components of the biomass for a very long time after their first emergence. But the biomass associated with these grasses may have quickly become important to hominins. Without exception, all δ¹³C studies of hominin diets have shown that significant proportions of their diets contained C₄-derived carbon (Lee-Thorp et al. 2003) (see Sponheimer and Lee-Thorp Vol. 1, Chapter 17). The implication is that when grass-related foods became available, hominins chose to make use of these new resources, a development that may be suggested as a defining characteristic of hominins not shared with our closest living relatives (Sponheimer et al. 2005). This hypothesis has yet to be tested on earlier material. On present evidence, it would seem that a link between hominin emergence and environmental change is still there, but it is not quite as we anticipated it would be.
We can also use the isotopic data to evaluate other hypotheses linking climate or environment and evolution. There seems to be no strong isotopic evidence to support the idea of a large climate-driven shift about 2.4–2.8 Ma, as suggested in the turnover-pulse hypothesis (Vrba 1980). Rather, the evidence from δ13C in fauna and pedogenic carbonates from East and South Africa overwhelmingly suggests that the most significant biomass change to more open, grassy ecosystems occurred near 1.8 Ma. All available evidence suggests that this was indeed a big change. It is tempting but nevertheless speculative to suggest that the rise and success of a habitual, fully bipedal, more technologically advanced hominin, viz. *Homo erectus sensu lato*, is in some way connected with this widespread shift in African environments.

The nature of much of the isotopic evidence produced to date does not permit a detailed test of the variability selection hypothesis (Potts 1996, 1998). This is because we tend to produce data averaged in large time chunks, and the variability is mostly hidden. The new speleothem isotope paleoenvironmental data (Hopley 2004), as well as data emerging from East African lakes (Trauth et al. 2005), represent new departures. The Buffalo Cave δ18O and δ13C data sequence suggests orbital control on rainfall and vegetation, respectively, for ~0.5 Ma in the Pliocene and Pleistocene. That the δ18O data show a dominant precessional cycle very similar to the cycles evident from the Tswaing Crater in the last 180 Ka (Partridge et al. 1997) suggest that it is a persistent feature. The difference in cyclicity between δ18O and δ13C underscores the complexity of climate control on environmental variables, where it can be discerned. Above all, and in agreement with the Saharan dust cycles (deMenocal 1995), the data suggest that the obliquity and precessional orbital cycles remained underlying drivers of environmental variability well before the advent of the very large 100 Ka climate swings after ~0.8 Ma. This kind of sequence may at last provide greater land-based substance for testing Potts (1996, 1998) variability selection hypothesis.

**References**


Dansgaard W (1964) Stable isotopes in precipitation. Tellus 16: 436–468


thousand years in southern Africa. Quat Sci Rev 22: 2311–2326
Lee-Thorp JA, Beaumont PB (1995) Vegetation and seasonality shifts during the late Quaternary deduced from $^{13}\text{C}/^{12}\text{C}$ ratios of grazers at Equus Cave, South Africa. Quat Res 43(3): 426–432


10 Chronometric Methods in Paleoanthropology

 Günther A. Wagner

Abstract

The aim of archeochronometry is the numeric dating, that is in term of years, of archeological and paleoanthropologic events or processes. The methods that are currently applied with most success are all based on the physical phenomenon of radioactivity. Their development underwent in the last few decades—and still undergoes—rapid progress. It is, in particular, the improvement in time resolution but also the application to novel sample materials as well as the extension of the age range of numeric dating that left a strong impact on modern paleoanthropology. This contribution introduces into the principles of radiometric dating. The most frequently applied dating methods, such as the potassium-argon, the uranium series, the fission track, the luminescence, the electron spin resonance, and radiocarbon techniques, are described. Their potential for paleoanthropology is illustrated using various examples covering the period since human entered the scene few million years ago.

10.1 Introduction

In paleoanthropology—as in any historically oriented discipline—time when something happened is the fundamental quantity. Recorded events do not make sense until they are arranged in their correct chronological order. Once such order is known, a mutual, causal relationship between the events may be confirmed or rejected. For instance, the coexistence of Neanderthals with modern humans excludes any hypothesis of simple evolution of the first into the second. Furthermore, a firm chronological database enables assessing the duration and the rate of processes, as is exemplified in the case of the remarkably fast-spreading of Homo erectus out of Africa. It is exactly for these reasons that the rapid progress, which chronometric dating experienced during the past few decades, inspired substantially the development of modern paleoanthropology.
Chronometry refers to the measurement of past time in terms of years, i.e., it leads to numeric age results. Before the introduction of chronometric dating to prehistory, age determination relied essentially on the principles of stratigraphy, which reveals merely the relative order of the records. The turning point came around 1950 with the development of radiocarbon dating, soon followed by potassium–argon and other dating techniques. By now, there are several chronometric dating methods available, which are based on physical phenomena. Since the result of a physical measurement represents only a more or less accurate approximation of the sought after, true value, which is and remains unknown, the often used term “absolute” dating seems inappropriate and should be avoided. Instead the term “chronometric” or “numeric” dating is recommended. The estimated deviation of the measured age value from the true one is given by the error. The error $\pm \sigma$ describes an interval from $(t - \sigma)$ to $(t + \sigma)$ surrounding the measured age $t$ so that the true value is expected with a certain probability within this confidence interval. Consequently, the error is an integral part of the age result and must not be omitted.

For numerical dating, one needs “clocks,” i.e., natural time-dependent processes that lead to quantifiable changes within prehistoric timescales, whereby the rate of the process needs to be known. Furthermore, the process must be reset—like a stopwatch—at the moment that is of interest to the paleoanthropologist, for example, in case of a stony artifact it is not the age of the stone itself but that of the manufacture which is of major concern. Processes which fulfill these requirements are rare. It was the discovery of radioactivity by Henri Becquerel in 1896 which provided such process. Because radioactive decay is a property of the atomic nucleus, its decay rate is insensitive to ambient parameters, such as temperature, pressure, and chemical bonding, i.e., the radioactive clock runs steadily, regardless of environmental influences. For this unique property, radioactivity may aptly be called the “mother of chronometry.”

A nuclide is an atomic species that is characterized by its numbers of nuclear protons and neutrons. Their sum is the mass number, which is indicated as superscript to the upper left of the element symbol. To give an example, the nuclide $^{14}$C consists of six protons, which define the element carbon, and eight neutrons. Nuclides belonging to the same element are called isotopes, i.e., they have the same number of protons but different mass numbers, such as the three carbon isotopes $^{12}$C, $^{13}$C, and $^{14}$C. Most natural nuclides are stable, but others disintegrate spontaneously—a phenomenon called radioactivity. The rate of this disintegration is nuclide-characteristic. The radioactivity $\frac{dN}{dt}$ (in the unit of becquerel, Bq [s$^{-1}$]) is defined as the fraction $dN$ of the radioactive parent nuclide $N$ (with initial amount $N_0$) that disintegrates within the time interval $dt$ into the radiogenic stable daughter nuclide $D (= N_0 - N)$. 

Chronometric methods in paleoanthropology
\[
\frac{dN}{dt} = -\lambda N = -\frac{\ln 2}{t_{1/2}} \cdot N
\]

(1)

whereby \( \lambda \) \([a^{-1}]\) is the decay constant and \( t_{1/2} \) \([a]\) the half-life \((= 0.693/\lambda)\).

By integration, one obtains the equation

\[
\frac{N}{N_0} = e^{-\lambda t}
\]

(2)

and substituting \((D + N)\) for \(N_0\)

\[
\frac{D}{N} = e^{\lambda t} - 1
\]

(3)

The age \( t \) \([a]\) can be derived, depending whether \(N_0\), \(N\), or \(D\) is known, according to one of the following equations:

\[
t = \frac{1}{\lambda} \ln \left( \frac{D}{N} + 1 \right)
\]

(4)

\[
t = \frac{1}{\lambda} \ln \left( \frac{N_0}{N} \right)
\]

(5)

The use of a radioactive system for age determination presupposes that neither the parent nor the daughter nuclides are lost or gained except through the decay process itself—a condition that is known as closed system.

There are several types of radioactive decay: \(\alpha\)-decay takes place under, the emission of an \(\alpha\)-particle, which is a \(^4\)He nucleus. During \(\beta\)-decay, the nucleus emits a \(\beta\)-particle, which is an electron whereby a neutron is converted into a proton. Electron capture occurs when the nucleus captures an extranuclear, orbiting electron from the innermost atomic shell (K-shell) whereby a nuclear proton in converted into a neutron. During spontaneous fission, the atomic nucleus splits into two heavy fragments and two or three neutrons. Some nuclides exhibit a dual decay mechanism, as \(\alpha\)-decay and spontaneous fission for \(^{238}\)U.

Radioactive decay is generally accompanied by the emission of energy discrete \(\gamma\)-rays—a type of electromagnetic radiation coming from the exited nucleus.

Instead of being stable, the daughter nuclide may be radioactive and disintegrate itself. Several such radioactive daughter nuclides following each other form a decay chain until finally a stable end product is reached. Most prominent for chronometric dating is the decay chain starting from \(^{238}\)U and ending at \(^{206}\)Pb involving several steps of \(\alpha\)- or \(\beta\)-disintegration. If the decay chain stays undisturbed, i.e., under closed-system conditions, a balance between production and decay of the interim members is gradually established. At this stage, which is called radioactive or secular equilibrium, all radioactive members \(N_1, N_2, N_3\), etc.
Assume equal radioactivity \( \frac{dN}{dt} \)

\[
\dot{\lambda}_1 N_1 = \dot{\lambda}_2 N_2 = \dot{\lambda}_3 N_3 \quad \text{K, etc.} \quad (6)
\]

In nature, various kinds of radioactive nuclides occur and many of them can be used for chronometry. Due to their origin, these nuclides can be divided into various groups: Primordial nuclides are still left over from the time of nucleosynthesis and, thus, are older than the formation of the earth (e.g., \(^{238}\text{U}\)). Radiogenic nuclides are produced by radioactive decay (e.g., \(^{230}\text{Th}\)). Cosmogenic nuclides are formed by the interaction of cosmic rays with the atmosphere and the earth’s surface (e.g., \(^{14}\text{C}\)). Anthropogenic nuclides are produced in nuclear plants and explosions (e.g., \(^{3}\text{H}\)). For the Quaternary period, there is a wide spectrum of dating methods available (Wagner 1998). Figure 10.1 presents an outline of the radiometric dating methods with sound paleoanthropologic potential.

### 10.2 Potassium–Argon

The K–Ar method covers the whole age range from the beginning of the solar system to the Holocene and has become undoubtedly one of the most important chrono-
metric dating tools since the first attempts were done by Smits and Gentner (1950). The term “potassium–argon” stands for several dating techniques. Apart from the conventional K–Ar technique (\(^{40}\text{Ar}/^{40}\text{K}\)), these are the argon–argon (\(^{40}\text{Ar}/^{39}\text{Ar}\)) and the argon–argon laser techniques. In particular, the latter one plays a crucial role in Plio-Pleistocene chronology whenever volcanic materials are involved. All techniques rest on the same phenomenon—the radioactive decay of the potassium isotope \(^{40}\text{K}\) to the argon isotope \(^{40}\text{Ar}\) so that the amount of accumulated radiogenic argon, in relation to the potassium content, becomes a measure for the age. The various techniques essentially differ in their analytical and gas extraction procedures.

Potassium is the eighth most common element in the earth’s crust and occurs widely in rock-forming minerals such as feldspar and hornblende. One of its three natural isotopes, \(^{40}\text{K}\) with 0.01167% isotopic abundance, is radioactive and disintegrates under electron capture with the decay constant \(\lambda_e\) of \(0.581 \times 10^{-10} \text{ a}^{-1}\) to the stable argon isotope \(^{40}\text{Ar}\). The long half-life of 11930 Ma implies that in the time span of a few million years ago only very low amounts of radiogenic \(^{40}\text{Ar}_{\text{rad}}\) are produced. Therefore, in Quaternary applications potassium-rich minerals, such as sanidine, are preferred. The ultrasensitive analytical technology developed lately allows determining ages as low as few thousand years. The K–Ar clock dates events involving the last complete degassing. In Plio-Pleistocene context, such events are essentially the solidification of volcanic minerals. An important material in this connection are the widespread tephra layers, for which K–Ar dating yields excellent tephrochronologic time marks, in particular, if the tephra are intercalated in early hominid-bearing sedimentary layers.

The K–Ar age \(t\) \(\text{[a]}\) is assessed under the assumptions that, first, no \(^{40}\text{Ar}_{\text{rad}}\) was present in the sample at the moment of the last complete degassing and, second, the subsequently produced \(^{40}\text{Ar}_{\text{rad}}\) remained quantitatively in the sample. Argon is analyzed by mass spectrometry and potassium by atomic absorption. This technique is called conventional K–Ar dating.

Apart from radiogenic \(^{40}\text{Ar}\), there are additional contaminant sources of this isotope. Samples always contain more or less atmospheric argon \(^{40}\text{Ar}_{\text{atm}}\), which must be subtracted from the total \(^{40}\text{Ar}\) in order to obtain the radiogenic fraction \(^{40}\text{Ar}_{\text{rad}}\). This correction is possible owing to the known isotopic abundance of atmospheric argon \((^{40}\text{Ar}/^{36}\text{Ar})_{\text{atm}} = 295.5\), requiring the measurement of the isotope \(^{36}\text{Ar}\) in addition to \(^{40}\text{Ar}\). Argon as a noble gas is chemically inert and thus, should be driven out of the crystal lattice during sufficient heating so that the K–Ar clock is reset. If the degassing is incomplete, the extraneous \(^{40}\text{Ar}\) component is left over, resulting in an overestimate of the K–Ar age. Such nonatmospheric argon contamination needs to be identified and overcome, which is achieved by the single-grain or isochron-dating techniques. The younger
the rock, the less abundant is the radiogenic $^{40}\text{Ar}$ and the larger is the fraction from contaminant argon sources. Therefore for the age determination of Quaternary samples, special K–Ar techniques were developed, which allow analyzing minute proportions of radiogenic argon.

After the resetting event, the K–Ar system must stay closed. Partial loss of argon results in an underestimate of the K–Ar-age value. The leaking of argon from minerals may happen continuously during weathering or episodically during thermal overprint. For unweathered Quaternary volcanites that cooled quickly after eruption to surface temperature and stayed cool afterward, argon loss rarely causes any problem. If present, argon loss can be recognized and corrected by the step heating $^{39}\text{Ar}–^{40}\text{Ar}$ plateau technique.

In the $^{39}\text{Ar}–^{40}\text{Ar}$ technique, potassium is determined through a nuclear reaction instead of chemical analysis. The sample is irradiated with fast neutrons, whereby the argon isotope $^{39}\text{Ar}$ is produced from the main potassium isotope $^{39}\text{K}$. The abundance of the artificial isotope $^{39}\text{Ar}$ is measured together with $^{40}\text{Ar}$ and $^{36}\text{Ar}$. The age is determined against a standard of known age, which is irradiated jointly with the sample to be dated. The fact that instead of explicit quantities only isotope ratios need to be measured improves the accuracy. A further advantage of this technique over the conventional one is the ability to recognize argon loss or extraneous argon. This is achieved by stepwise heating and differential degassing the sample whereby for each step the $^{40}\text{Ar}/^{39}\text{Ar}$ ratio of the released argon is measured. In the absence of argon loss and of excess argon, the corresponding ages display an age plateau.

The laser technique is a variant of the $^{39}\text{Ar}–^{40}\text{Ar}$ technique. The heating is achieved under a laser beam. The sample is either incrementally heated and finally fused—analogous to the plateau technique—or without gradual degassing directly fused and analyzed. The main advantage of this technique is its ability of analyzing single grains in the submilligram range. The grain-discrete probing enables to identify contaminating detrital grain populations by their higher age.

To test whether the K–Ar system is disturbed by excess argon, the isochron technique is also used. It relies on cogenetic fractions from the same rock sample that have different potassium contents. In the isochron diagram, $^{40}\text{Ar}/^{36}\text{Ar}$ is plotted against $^{39}\text{Ar}/^{36}\text{Ar}$. All data points of the subsamples lie on a straight line (isochron) whose slope defines the age and whose intercept with the y-axis reveals the initial $^{40}\text{Ar}/^{36}\text{Ar}$ ratio: if no excess argon is present at the atmospheric ratio of 295.5 and if present at a higher value. Also $^{36}\text{Ar}/^{40}\text{Ar}$ versus $^{39}\text{Ar}/^{40}\text{Ar}$ diagrams are used. In this case, straight lines of negative slope define undisturbed systems. Usually isochron plots are applied in connection with argon–argon laser single-grain analysis. In such case, it is important to probe a sufficiently large number of crystals in order to discriminate between xenocrysts (older crystals incorporated
into the volcanic rock), phenocrysts (crystals that solidified from the volcanic melt before eruption), and the crystals formed during the volcanic eruption (Chen et al. 1996).

Of particular significance are tephra horizons that are intercalated in sedimentary sequences. The ages yield invaluable tephrochronologic and stratigraphic time markers. Our knowledge about the early hominids, their evolution, and dispersion rests essentially on K–Ar data of such tephra. However, tephra layers in sediments may be reworked and thus contaminated with older mineral detritus of various provenances. In such case, bulk K–Ar or Ar–Ar dating of mineral concentrates yields an integrated age that is too old with respect to the event of volcanic eruption. For this reason, but also for the recognition as well as correction of excess argon, single crystal probing is required in order to identify the various components of different age. This is optimally achieved with the grain-discrete Ar–Ar laser technique. Suitable mineral phases are potassium-bearing feldspars, such as sanidine and plagioclase, but also biotite, hornblende, and acidic glass shards.

One of the most renowned sites with early hominid fossils is Olduvai Gorge, Tanzania. The ca. 100 m thick Plio-Pleistocene sediments contain numerous tephra horizons and lava flows. Remains of *Australopithecus boisei* were uncovered as well as stone tools in Bed I. Continuous efforts to date this bed had aroused controversial views on *H. habilis* and its age. Finally, single grain $^{40}$Ar/$^{39}$Ar laser-fusion dating succeeded in establishing a detailed and reliable chronology (Walter et al. 1991). Feldspar grain populations of different ages were observed so that the decisive tuff component could be distinguished from older, reworked contamination. The weighted means of single grains from this juvenile component are $1.798 \pm 0.004$ for Tuff IB and $1.779 \pm 0.007$ Ma for Tuff IF within sedimentary layers that include the fossil hominids. The oldest Oldowan-type stone tools, discovered at Gona, Ethiopia, have an age of 2.6–2.5 Ma. They are associated with cut-marked bones. The dating of the artifact-bearing layer is based on an overlying tuff, which yielded $2.53 \pm 0.15$ Ma as weighted mean of single-plagioclase $^{40}$Ar/$^{39}$Ar ages, and the underlying Gauss-Matuyama paleomagnetic boundary, which is placed at 2.58 Ma (Semaw et al. 2003). One of the first attempts to date basalt flow with K–Ar is that by von Koenigswald et al. (1961) who studied the massive flow underlying Bed I at Olduvai and found an age of $1.3 \pm 0.1$ Ma, which is an underestimate probably due to argon loss. At the Pliocene site Fejej, Ethiopia, with dental remains attributed to *Australopithecus afarensis*, the fossil-bearing sandstone is capped by fine-grained basalt flows. The whole rock samples from two lowest basalt outcrops were analyzed with the laser $^{40}$Ar–$^{39}$Ar incremental heating technique. Both samples showed long age plateaus from which the means of $3.94 \pm 0.05$ and $4.06 \pm 0.07$ Ma, respectively, were
calculated. When combined with paleomagnetic data from the site a minimum age of 4.00–4.18 Ma was assigned to the fossil teeth, making them among the very oldest remains of *A. afarensis* (Kappelman et al. 1996).

As to the question when and how *H. erectus* migrated out of Africa, K–Ar data are of special interest. The earliest fossil trace of this species in Africa occurs in the Koobi Fora region, Kenya, at 1.8 Ma (McDougall 1985). According to K–Ar evidence, *H. erectus* seems to have appeared practically at the same time already in Western Asia. At Dmanisi, Georgia, fluvio-lacustrine sands with several hominid remains assigned to this archaic species are directly under- as well as overlain by volcanic flows and ashes. This stratigraphic situation allows the application of 40Ar–39Ar–age bracketing, resulting in 1.85 ± 0.1 and 1.81 ± 0.05 Ma for the lower and upper volcanic layers, respectively (de Lumley et al. 2002). Based on 40Ar/39Ar laser incremental ages (1.81 ± 0.04 and 1.66 ± 0.04 Ma) on hornblende, separated from pumice of two *Pithecanthropus* sites in Java, Swisher et al. (1994) claimed that *H. erectus* appeared concurrently in Southeast Asia. However, the stratigraphic position of the former hominid finds seems to be above the dated pumice, so that these 40Ar–39Ar ages can be considered merely as a *terminus post quem* for the arrival of the first hominids in Java (Semah et al. 2000).

For the emergence of anatomically modern humans, the recently discovered fossil *Homo sapiens* at Herto, Ethiopia, in fluvial and lacustrine sandstone with Lower and Middle Stone Age technocomplexes are of prime importance (Clark et al. 2003). The upper age limit for this sandstone unit was determined by 40Ar/39Ar incrementally heated multigrain analysis of anorthoclase from embedded pumice, yielding 163 ± 3 and 162 ± 3 ka, respectively, and of embedded obsidian clasts, yielding 160 ± 2 ka. The lower age is constrained by the 40Ar/39Ar laser fusion mean age of 154 ± 7 ka of single grains from the juvenile sanidine-component, separated from a tuff which caps the fossiliferous unit. These data provide fresh evidence for the out-of-Africa model of *H. sapiens*’ origin.

### 10.3 Uranium series

The general term “Uranium series” comprises several closely related dating methods, which are based on the radiometric disequilibrium within the radioactive decay series arising from the two uranium isotopes 238U and 235U (Ivanovich and Harmon 1992). Also the terms decay series, disequilibrium, or uranium–thorium methods of dating are occasionally applied. In paleoanthropology, it is essentially the 230Th/234U method, which is most important for calcareous remains up to about 400 ka old. Of particular interest for 230Th–234U dating
are secondary carbonates of caves and springs and to a lesser degree also fossil teeth and bones. Occasionally also the $^{231}$Pa/$^{235}$U method is applied up to 150 ka.

The uranium isotope $^{238}$U, which constitutes 99.3% of natural uranium, is radioactive and decays over a chain of intermediate radioactive daughter nuclides to the stable lead isotope $^{206}$Pb. In closed systems, equilibrium develops among all radioactive nuclides within the decay chain. In the state of radioactive equilibrium, all radioactive nuclides possess equal activity (Eq. 6). In nature, most minerals and unweathered rocks represent closed systems, in which radioactive equilibrium persists simply as a result of their geologically high ages. If such a system is disturbed it will take sometime, practically five half-lifes, until the daughter nearly returns to equilibrium with its parent nuclide. For the system $^{230}$Th/$^{234}$U, in which $^{234}$U decays by $\alpha$-emission to $^{230}$Th, disequilibrium arises through geochemical fractionation. Uranium is readily dissolved in groundwater from where it is taken up by secondary carbonates as well as dental/bone tissues. Thorium, on the other hand, stays adsorbed to mineral particles and thus is not dissolved in groundwater. Consequently, fresh calcareous deposits incorporate uranium but not thorium, i.e., $(^{230}$Th/$^{234}$U)$_{t=0} = 0$. During the following 400 ka, corresponding roughly to five times the $^{230}$Th half-life of 75.6 ka, the $^{230}$Th/$^{234}$U system gradually builds up to equilibrium. The time-dependent increase of $^{230}$Th/$^{234}$U enables the determination of the time $t$ elapsed since the event of disturbance.

The essential requirements for dating are that the initial abundance of the daughter is, unless known, negligibly small and that after the disturbance the radioactive system remains closed. Complications in $^{230}$Th–$^{234}$U dating have various sources: (1) During the incorporation, $^{234}$U is not necessarily in equilibrium with its radioactive predecessor $^{238}$U, since in groundwater the activity ratio is $^{234}$U/$^{238}$U > 1 so that the $^{234}$U/$^{238}$U ratio must be determined and taken into consideration for age calculation. (2) The requirement that at deposition the sediment is free of $^{230}$Th is frequently not fulfilled since thorium, and consequently also $^{230}$Th, may be present. Negligible $^{230}$Th contamination is indicated by $^{230}$Th/$^{232}$Th > 20. (3) The system may experience a secondary opening as a result the geochemical mobility of the uranium.

Originally, the activity of the Th and U isotopes was analyzed by $\alpha$-spectrometry. The introduction of thermoionization mass spectrometric (TIMS) by Edwards et al. (1986/87), which requires less than 1 g of sample material and provides high age precision, brought a great impetus for $^{230}$Th–$^{234}$U dating. Under favorable circumstances age precision better than 1% can be obtained. The uranium content should be more than 0.1 $\mu$g/g.

When sampling speleothems (calcareous flowstones formed on the cave floor), dense and pure carbonates should be collected in order to minimize the
dangers of open system behavior and $^{230}\text{Th}$ contamination by detrital components. In the Tongtianyan Cave, Guangxi, China, one of the few fossils of *H. sapiens* in China was discovered. Several speleothem layers, intercalated in the cave sediments, were dated by $^{230}\text{Th}/^{234}\text{U}$. The age results of $13.5 \pm 2.3$, $63.1 \pm 2$, and $148 \pm 4$ ka for three layers from top to bottom, respectively, follow the stratigraphic order (Shen et al. 2002). Although the stratigraphic position of the hominid find is to some extent uncertain, it is certainly below the second layer. This renders this finding as one of the earliest in East Asia, indicating that the Moderns arrived there before 63 ka. Depending on the measurement precision, the $^{230}\text{Th}/^{234}\text{U}$ system in speleothems older than 350 ka often shows equilibrium so that only a *terminus ante quem* can be given for the age. This situation was met in the fossil hominid sites known as the Sima de los Huesos, Spain, and the Caune de l’Arago, France. At Sima de los Huesos, the numerous human individuals, which are considered as evolutionary ancestors of the Neanderthal, are overlain by a speleothem that is in equilibrium and thus older than 350 ka (Bischoff et al. 2002). Analogously, the pre-Neanderthals found in the Middle Pleistocene unit III at Caune de l’Arago must be older than 350 ka (Falguères et al. 2004). At the famous Zhoukoudian site near Peking, with the occurrence of *H. erectus*, commonly known as Peking Man, speleothems that are intercalated in the fossiliferous sediments were $^{230}\text{Th}/^{234}\text{U}$-dated. In most samples, the $^{230}\text{Th}/^{234}\text{U}$ system was close to equilibrium. However, high-precision analysis of these nuclides enabled to push the dating method to its upper limits, which is at 500–600 ka age range with the present instrumentation (Shen et al. 2001). The results indicate that the youngest member of *H. erectus* at this site is >400 ka and the oldest members from the lower layers significantly older than 500 ka.

The $^{230}\text{Th}/^{232}\text{Th}–^{234}\text{U}/^{232}\text{Th}$-isochron technique permits determining the initial $^{230}\text{Th}/^{232}\text{Th}$ ratio. Schwarcz (1989) used this technique for the dating of the sinter crust on the cranium of the classic Neanderthal at Monte Circeo, central Italy. The encrustation on the cranium consisted of a brighter inner and a dark-brown outer layer. The outer layer provided 16 ka. From the inner layer, containing detrital contamination, several subsamples of different U/Th ratio were obtained through fractionated leaching. The slope of the straight line (isochron) through the data points in the $^{230}\text{Th}/^{232}\text{Th}–^{234}\text{U}/^{232}\text{Th}$ diagram gives an age of $51 \pm 3$ ka.

Precise $^{230}\text{Th}–^{234}\text{U}$ dating of travertine, a calcareous spring-sinter, is hampered by the common detrital $^{230}\text{Th}$ contamination and by open system behavior. Applying a microsampling technique, in which 100 mg were selectively drilled from the micrite/spar phases, Mallick and Frank (2002) dated successfully travertine from various Thuringian sites, among them Weimar–Ehringsdorf with pre-
Neanderthal remnants. The results assign these finds firmly to oxygen isotope stage 9.

Buried teeth absorb uranium from the groundwater. The knowledge of the time–function of the uranium uptake is crucial for the $^{230}\text{Th}/^{234}\text{U}$-age evaluation. Since the exact temporal development of uptake is unknown, one has to rely on models such as early uptake (EU) or linear uptake (LU) (cf. ESR). For instance, in order to strengthen dating of the long coexistence of early Moderns and Neanderthals in the Levantine corridor (cf. luminescence), McDermott et al. (1993) applied mass spectrometric $^{230}\text{Th}–^{234}\text{U}$ dating to mammal teeth from the Middle Pleistocene sites Tabûn, Qafzeh and Skhûl, Israel. They showed that at Tabûn, both types of humans were approximately coeval some $100 \pm 5\text{ ka}$. The analyses were carried out on 50–100 mg subsamples of dentine and enamel of the same tooth, whereby it is important that the ages agree to each other, although both subsamples differ strongly in U content and in porosity, indicating EU and closed-system behavior. The mode of uranium uptake can be constrained when $^{230}\text{Th} / ^{234}\text{U}$ is combined with ESR-dating (Grün et al. 1988), which in the meantime became the common practice. Falguères et al. (1997) reported coupled $^{230}\text{Th} / ^{234}\text{U}$ and ESR-dating on horse teeth from Acheulian and Mousterian levels at Micoque, France. The $^{230}\text{Th} / ^{234}\text{U}$ ages of enamel and dentine range widely from 150 to $>350\text{ ka}$. However, when combined with ESR-dating, consistent ages between 300 and 350 were obtained.

10.4 Fission track

Although fission tracks (FT) are not applied as commonly as the other radiometric dating methods in paleoanthropology, they made significant contributions at some important sites in volcanic regions. They are formed by the spontaneous nuclear fission of uranium. Natural uranium consists of the isotopes $^{238}\text{U}$ (99.3%) and $^{235}\text{U}$ (0.7%), whereby $^{238}\text{U}$ decays by spontaneous fission. The decay rate of this fission is $10^6$ times less than that of the $\alpha$-decay of the same isotope. During fission, the uranium nucleus splits up into two fragments. Due to their kinetic energy, both fission fragments are expelled in opposite directions and leave along their path a zone of radiation damage. Both branches together form a straight fission track of $10–20\text{ \mu m}$ in length and several $10^{-3}\text{ \mu m}$ in diameter. By chemical etching, the fission tracks can be made visible under the optical microscope. In the course of time, the tracks accumulate in the minerals. When all tracks are preserved, their number gives the age of the sample. Since the track number depends also on the uranium content, the latter one needs to be known. For U analysis, the thermal-neutron induced fission of $^{235}\text{U}$ is exploited.
The number of the induced $^{235}$U fission tracks is proportional to the U content. Thus, the procedure of fission-track dating essentially involves the counting of spontaneous $^{238}$U fission tracks before and induced $^{235}$U fission tracks after a neutron irradiation.

The principles and application of fission-track dating were described in detail by Wagner and Van den haute (1992). With fission tracks, one dates either the formation or a secondary heating event at which all previous tracks were erased, i.e., the clock was reset. The fission-track method is applicable to ages of $>10$ ka. This requires, however, sufficiently high uranium contents above 100 $\mu$g/g. Zircon, due to its high uranium content, is most frequently used in the paleoanthropologic age range. This mineral may occur in volcanic rocks. Of particular interest are volcanic ashes that are intercalated in sedimentary sequences containing hominid remains and Paleolithic implements. Also volcanic glass, such as obsidian and pumice, is frequently used for fission-track dating.

A commonly met problem in fission-track dating is track annealing. Latent fission tracks gradually fade over time. The fading is accelerated at elevated temperatures, a process known as annealing. Since annealing reduces the apparent fission-track age, it is of fundamental importance. Fading can be recognized by track-length measurements since annealing shortens the tracks. Fortunately, tracks in zircon are rather stable and do not show any signs of fading over several million years at ambient temperatures, although tracks in natural glasses certainly may fade under such conditions.

For fission-track dating of tephra, mainly zircon grains and, to a lesser degree, also glass shards and apatite as well as titanite grains are used. When relying on heavy minerals, the problem of different provenance of the various grains needs to be taken into consideration, a difficulty already discussed (cf. K–Ar dating). Primary volcanic grains in presence of detrital ones can be identified—apart from mineralogical criteria—by single-grain fission-track data. A good case study is that on the Plio-Pleistocene sedimentary sequence of the Koobi Fora formation, Kenya. It contains several tuff horizons, which primarily consist of glass fragments and pumice cobbles and show signs of redeposition. Of particular interest is the KBS Tuff, which is intercalated in hominid-bearing layers. K–Ar data on the KBS Tuff raised in the 1970s a controversy between supporters of a long chronology ($2.61 \pm 0.26$ Ma) (Fitch and Miller 1970) and those of a short chronology ($1.82 \pm 0.04$ Ma) (Curtis et al. 1975). FT dating on zircon ($2.44 \pm 0.08$ Ma) (Hurford et al. 1976) at first seemed to support the high K–Ar age. A later FT study of zircon from the pumice yielded $1.87 \pm 0.04$ Ma (Gleadow 1980) in accordance with the low chronology. Besides methodological aspects, the main reasons for the previous fission-track overestimate of the KBS Tuff are detrital, old zircon grains. A far-reaching study on tuffaceous zircon was
reported by Morwood et al. (1998). At the site Mata Menge, Flores, Indonesia, a layer with stone tools is intercalated in tuffaceous layers. FT dating on zircon from the lower and the upper layer yielded \(880 \pm 70\) and \(800 \pm 70\) ka, respectively. Provided that these grains are primary and not reworked, these findings imply that at that time \(H. \text{ erectus}\) had already reached the island of Flores from Southeast Asia—a journey that requires an amazing sea-crossing capability, even at periods of lowest sea level.

Ashes at prehistoric fireplaces may contain sufficiently heated grains of apatite, zircon, and titanite. Such case was encountered at Zhoukoudian near Peking, with its numerous remains of \(H. \text{ erectus}\), the Peking man. From ashes of the layers 10 and 4, several hundred grains of titanite in the size range of 50–300 \(\mu\)m were separated. As criterion for discriminating completely from partially annealed titanite grains, the length of the fission tracks was utilized. Altogether, 100 grains showed complete resetting and gave mean ages of \(462 \pm 45\) ka for layer 10 and \(306 \pm 56\) ka for layer 4 (Guo et al. 1991), being significantly less than the already mentioned uranium series ages for this site (Shen et al. 2001).

### 10.5 Luminescence

Since its introduction by Daniels et al. (1953) luminescence dating has gradually developed into a powerful chronometric technique, particularly for quartz- and feldspar-bearing materials (Aitken 1985, 1998). In the meantime, luminescence dating has significantly contributed to paleoanthropology. As to the techniques of luminescence dating, one distinguishes between thermoluminescence (TL) and optically stimulated luminescence (OSL). For the latter, the term “optical dating” is also used. Luminescence dating covers a wide age range between 10 and \(10^5\) years and thus is able to reach well beyond the limits of radiocarbon dating. Datable materials comprise various inorganic sediments, such as sand and loess, heated stones, and bleached stone surfaces.

Luminescence dating is based on the time-dependent deposition of energy in the crystal lattice of minerals. This energy stems from ionizing radiation, which originates from natural radioactivity as well as cosmic radiation and is omnipresent in nature. When this radiation, consisting of energetic \(\alpha\) and \(\beta\)-particles as well as photons (\(\gamma\)-rays), interacts with the atoms of a mineral, it transfers energy to them so that electrons can be removed from their original valence-band position in the atom shell. Freed electrons diffuse for a short distance through the crystal lattice and some of them become trapped at lattice imperfections. Trapped electrons are at higher energy levels than those in the valence band. With time \(t\), the electron traps are increasingly filled—the process, which forms the
basis of the luminescence clock. When the crystal is stimulated by heat or light, the electrons are released from their traps, enabling them to recombine with opposite charge carriers whereby the formerly trapped energy is set free. Some of this energy appears as emission of visible light, the luminescence. Depending on the kind of stimulation, one differentiates thermally (TL) from optically stimulated luminescence (OSL). OSL is further split up according to the type of stimulating light such as GR-OSL by green or IR-OSL by infrared stimulation. The intensity of the luminescence signal is related to the accumulated energy dose AD (in unit of Gray, Gy) and, thus, to the time interval ("age" $t$) during which the mineral has been exposed to ionizing radiation. The luminescence age is calculated from AD and the dose rate DR (dose per time, in unit of Gya$^{-1}$).

$$t = \frac{AD}{DR}$$ (7)

From the above equation, it becomes clear that the dating procedure consists of two steps: the determination of AD and DR.

The growth of the luminescence signal with radiation dose starts linearly. However, with increasing dose the number of empty traps that are still available becomes fewer so that the growth curve assumes the shape of exponential saturation. In most cases, saturation is reached after doses of few $10^2$ Gy. This behavior restricts luminescence dating to the last few $10^5$ years. In order to convert the luminescence signal into a dose value, one has to know the sensitivity $S$ (luminescence signal per dose, i.e., the slope of the growth curve). It varies from sample to sample. For a given sample, it is the same for $\beta$- and $\gamma$-radiation but different for $\alpha$-radiation. The sensitivity ratio $S_\alpha/S_\beta$, the so-called $a$-value, needs to be determined but is mostly around 0.1.

Another behavior that limits the age range is fading of the latent luminescence signal in the course of time, i.e., the prerequisite that all centers involved in the signal generation are stable over the complete age range in question. Like any other type of radiation damage, latent luminescence signals are subjected to fading whose kinetics is essentially thermally controlled. As far as near-surface materials at normal ambient temperatures are concerned, natural fading limits the datable age range up to a few $10^5$ years.

An important concept in dating is the resetting of the system: the luminescence systems need to have been reset at the event that is of interest. Complete or at least partial resetting of the latent luminescence signals is caused by exposure to heat or light. Consequently, the last occurrence of such events can be dated, for example, the deposition of sediments or the heating of flint artifacts.

The past few years have seen a lot of progress in the laboratory protocols. Previously it has been the common practice to prepare multiple subsamples (aliquots) and apply various doses in addition to their natural one in order to
evaluate AD. Lately, the so-called SAR-protocol (single aliquot regeneration) is increasingly preferred which regenerates the luminescence signal after artificial resetting (Murray and Wintle 2000). When using regenerated growth curves, AD is evaluated through the value of the artificial dose that exactly matches the intensity of the natural one. In order to correct for sensitivity changes due to the laboratory procedure, a test dose is applied to the aliquot. The advantages of the SAR protocol over the conventional multiple aliquot technique are smaller sample size, less time for sample preparation, and improved analytical precision due to replicate AD determination. The present technology is directed toward single-grain protocols which allow making sure that all grains of a sample show the same luminescence age, which bears a great potential for novel applications (Roberts et al. 1997; Greilich et al. 2002; Jacobs et al. 2003).

Apart from the dose AD, the natural dose rate DR needs to be determined for the age calculation. The ionizing radiation at the earth’s surface originates predominantly from the radionuclides $^{232}$Th, $^{235}$U, $^{238}$U, and their daughter products as well as $^{40}$K and $^{87}$Rb and to a minor extent from cosmic rays. These nuclides emit $\alpha$, $\beta$, and $\gamma$-radiation, each of which has a different penetration depth, which amounts in rocks to ca. 20 $\mu$m, 1 mm, and 30 cm, respectively. One needs to take into account the internal component originating within the luminescence sample as well as the external one from the immediate surroundings. The age determination requires materials of a uniform and defined dose rate. For this reason, one separates the sample into certain sizes, usually the fine grain (4–10 $\mu$m) or coarse grain (100–200 $\mu$m) fractions. Since any water residing in the pore volume of the sample attenuates the dose rate, the moisture content and its possible temporal variation in the sample as well as its environment need to be known. Also the on-site intensity of the cosmic rays has to be assessed. It increases with topographic altitude and decreases with depth below the surface. Temporal variation of the dose rate may be caused also by changing contents of radionuclides due to disequilibrium within the decay chains. All these complications must be considered carefully since they may cause major uncertainties in luminescence dating. For dose-rate evaluation, several techniques are available, which include $\alpha$- and $\beta$-counting, $\gamma$-spectrometry, atomic absorption, and neutron activation analyses.

Luminescence dating of clastic sediments enables to determine when they were formed provided all of the sedimentary grains were sufficiently exposed to daylight before and during deposition. In this way also paleoanthropologic remains can be dated that are embedded in such sediment series. In Central Europe—as well as in other periglacial areas—numerous loess profiles with Paleolithic finds have been TL-dated (Zöller et al. 1991). Also many sands have been dated by OSL. An example is the dune sands of the Acheulian open-air site
of Holon/Israel where alkali feldspar and quartz fractions were dated by OSL as well as TL (Porat et al. 1999). However, the presumption that all grains had been completely bleached at deposition is not necessarily fulfilled. In particular fluvial sands, whose grains were transported under water cover, may contain partially bleached or even unbleached grains. In this case, the apparent luminescence age would be an overestimate. Also postdepositional vertical mixing between sedimentary layers, such as bioturbation, leads to erroneous ages. Luminescence dating of individual grains is able to identify these disturbances. This potential was convincingly demonstrated in the case of the Jinmium rock shelter in northern Australia. Fullagar et al. (1996) reported for the artifact-bearing floor-deposit TL ages of 176 ± 16 and 116 ± 12 ka, predating the first arrival of humans in Australia by more than 100 ka. The data were determined on multiple grain aliquots of quartz. Most scholars received these ages with skepticism and as a possible cause for the too high age, contamination with insufficiently bleached grains was assumed. An intensive OSL-dating program on the same deposits by Roberts et al. (1999), using the single-grain approach, yielded ages less than 10 ka when considering only those grains which had been fully bleached before burial. Single-grain OSL ages of quartz from Malakunanja II date the early human occupation of this Australian site to 55.5 ± 8.2 ka (Roberts et al. 1998).

**Burned flint** is well-suited for TL dating. Due to a relatively low internal dose rate and good TL stability behavior, its datable age range reaches back to about 500 ka and thus covers all phases of the Paleolithic. Clearly, TL dating of flint requires prehistoric annealing (>450°C) of the TL signal, which, fortunately, has been the case for a considerable number of flint artifacts. It is advisable to collect several flint samples from each layer to be dated. Intensive TL-dating studies were carried out at several Levantine sites with rich Lower to Middle Paleolithic lithic industries and human remains of Neanderthals as well as early Moderns. From the site of Tabûn/Israel, Mercier and Valladas (2003) reported stratigraphically consistent TL ages between 302 ± 27 and 165 ± 16 ka, indicating that the technological transition from the Acheulian to the Mousterian occurred some 250 ka ago. At the nearby site of Kebara, a skeleton of a Neanderthal and at the sites of Qafzeh and Skhûl remains of a Proto-Cro-Magnon, a predecessor of modern human, had been unearthed. TL dating on 20 flint fragments from the hominid-containing layer in Qafzeh yielded 92 ± 5 ka. At Kebara, 30 flint artifacts from several layers provided TL ages from 50 to 70 ka, and the layer with skeleton of the Neanderthal gave 60 ka. At Skhûl six burnt flints from the level with the remains of a Proto-Cro-Magnon yielded a TL age of 119 ± 18 ka (Mercier et al. 1993). These data reveal that early forms of modern human existed significantly earlier and synchronously with Neanderthals. At the early Upper Paleolithic site of Geissenklösterle/Germany, seven flint artifacts from the Proto-Aurignacian and
two from the Aurignacian levels ultimately satisfied the criteria of being sufficiently heated (Richter et al. 2000). The samples from the Proto-Aurignacian layer gave a weighted mean of 40.2 ± 1.5 ka. The mean TL age of the overlying Aurignacian level is 37.0 ± 1.4 ka. Altogether these data imply a much earlier beginning of the Upper Paleolithic in central Europe than anywhere in western and southwestern Europe.

A novel development is the OSL dating of stone surfaces that in the past have been exposed to daylight, with resetting of the OSL signal, and have been shielded from light since then. In such case, the moment of the last exposure to daylight is determined. This approach opens—at least in principle—a large potential for archeologic, paleoanthropologic, and geomorphologic applications such as the dating of stone structures and buried lithic implements as well as deposited boulders. Hitherto, the new method has been corroborated for granitoid rocks and successfully applied, in particular to the famous Nasca geoglyphs in southern Peru (Greilich et al. 2005).

### 10.6 Electron spin resonance

Electron spin resonance (ESR) dating is also based on the accumulation of radiation-induced energy in minerals and thus has close links to luminescence dating. Although first attempts to exploit the ESR phenomenon for dating go back to the 1960s (Zeller et al. 1967), as a dating method it did not flourish before the 1980s and is still being developed. The ESR method permits age determination up to few million years, far beyond the range of the luminescence methods, and covers the whole Quaternary period. The most important material for paleoanthropologic ESR application is tooth enamel, but quartz-separates from sediments at prehistoric sites also have a certain potential (Rink 1997). With ESR, one dates either the burial, as in the case of fossil teeth, or the resetting of a previous system as a result of bleaching, as in the case of sedimentary quartz grains or heating of stones.

The ESR phenomenon is caused by paramagnetic centers in the crystal lattice. Radiation-induced trapped electrons, mentioned already in the context of luminescence dating, form such centers and give rise to characteristic ESR signals. The intensity of the ESR signal is a function of the number of trapped electrons and, therefore, of the accumulated energy dose $AD$ that has been absorbed from the ionizing radiation in the course of time. In order to calculate the ESR age, the value of $AD$, obtained from the ESR measurement, is divided by the dose rate $DR$, in the same manner as already discussed for luminescence (Eq. 7).
The quantity of AD is determined by ESR spectrometry, which exploits the fact that trapped electrons are unpaired. Brought into a variable magnetic field and exposed to a given microwave, unpaired electrons show spin resonance at a specific strength of the magnetic field. The condition at which resonance happens is described by the $g$-value, which is characteristic for the type of the paramagnetic center. The energy necessary for the resonance is absorbed from the microwave so that its intensity reduction is a measure for the concentration of the center. The resulting ESR spectrum shows the specific microwave absorption for various centers with different $g$-values. Owing to measurement–technical reasons, the ESR spectra of the microwave absorption are not directly recorded; instead their first derivation as a function of the field strength is plotted. For the evaluation of AD, known doses are applied additively to the sample and a growth curve is established. ESR has the advantage over luminescence that the concentration of the probed centers is not disturbed by the measurement procedure, thus permitting one to establish the growth curve on the same aliquot. Most samples show exponential saturation functions. At normal ambient temperatures of sediments, most ESR centers are sufficiently stable for applying ESR dating up to few million years. The dose-rate determination follows the same principles as already mentioned for luminescence. For ESR-dating samples of a few grams are sufficient, provided the material is homogeneous, but usually larger sample sizes are preferable as they allow one to separate suitable materials in the laboratory and to conduct microdosimetric measurements.

ESR dating of fossil *tooth enamel* plays an important role for Paleolithic sites, mainly because teeth are commonly preserved. Furthermore, the age range covered by ESR dating reaches beyond those of radiocarbon or luminescence dating so that even the Lower Paleolithic becomes accessible. Mammalian dental issue consists essentially of enamel, dentine, and cementum layers. ESR dating is based on the mineral hydroxyapatite, in particular on its carbonate-containing subspecies dahllite, within the enamel. Dentine and cementum are less dense, contain more organic tissue, and take up uranium more easily than dentin. For these reasons, they are not used for ESR analysis but must be taken into account for microdosimetric reasons. The ESR spectrum of tooth enamel has, at $g = 2.0018$, a suitable signal of good sensitivity and high thermal stability.

The quantity of the dose rate introduces considerable problems into the ESR age evaluation. In teeth, one observes often strongly varying uranium contents on microscopic scale causing steep dose-rate gradients, which is in particular the case for the $\beta$-component since the $\beta$-radiation range (ca. 1 mm) is similar to the thickness of the enamel layers. Teeth gradually take up uranium from the groundwater after burial. *In vivo* they contain less than 1 $\mu$g/g U, but fossil ones up to 1,500 $\mu$g/g. This means that the dose rate increases with time and that the
closed-system condition is not fulfilled. To allow for time dependence of the dose rate, distinct models of uranium uptake are assumed such as the early uptake (EU), the linear uptake (LU), and recent uptake (RU) models. The EU-model results in a lower ESR age compared to the LU model due to a higher dose rate on the average. Both model ages may considerably differ from each other, especially for high uranium contents so that samples low in uranium are preferable. Most published ESR-age data are based on assumed U-uptake models. In order to set constraints to the validity of the model, ESR dating is coupled with uranium series dating (Grüner et al. 1988). The comparison the closed-system ages of $^{234}\text{U}/^{238}\text{U}$ and $^{230}\text{Th}/^{234}\text{U}$ with that of ESR enables to discriminate among the hypothetical models. Combined ESR/uranium series dating has become more and more the routine for tooth enamel dating.

The Early and Middle Paleolithic sites of Tabûn, Kebara, Skhûl, Qafzeh and Hayonim, Israel, with rich lithic and human bone inventories play a fundamental role with respect to the early human out-of-Africa dispersion through the Levantine corridor. In order to establish a firm chronology of this migration, ESR-dating has been repeatedly applied to mammalian tooth enamel from these sites, in addition to uranium series and luminescence. Worth mentioning in this connection are, in particular, two results. First, the combined ESR/uranium series tooth age of $387 \pm 50$ ka confirms a $340 \pm 33$ ka TL age on burnt flint for the Lower Paleolithic Yabrudian lithic industry at Tabûn, whereby the U-uptake appears to be more recent than linear (Rink et al. 2004). Second, the ESR ages on teeth from Middle Paleolithic contexts support the early dating of the anatomically Moderns to $80–120$ ka, with the amazing consequence that in the Levant the Moderns and Neanderthals coexisted for about $60$ ka (Grüner and Stringer 1991).

At the well-known South African site Swartkrans, with its wealth of remains of *Australopithecus robustus*, Curnoe et al. (2001) determined one of the oldest ESR ages so far reported. Coupled ESR/uranium series dating on two human and two bovid teeth yielded $1630 \pm 160$ ka with a possible maximum of $2110 \pm 210$ ka, indicating that ESR dating can provide reasonable results for samples of Late Pleistocene/Early Pleistocene age. Also the Atapuerca Gran Dolina site, Northern Spain, with the earliest humans in Europe, *H. antecessor*, was investigated by combined ESR/uranium series dating. From the bottom of Aurora stratum, where the very early human remains had been recovered, three ungulate teeth gave a mean of $731 \pm 63$ ka, which is in agreement with the paleomagnetic age estimate of $>780$ ka (Falguèrèrs et al. 1999).

ESR also permits the dating of calcareous sinter deposits, which show complex ESR spectra partly attributed to organic radicals. Important Paleolithic cave sites have been dated, among them Caune de l’Arago at Tautavel, southern
France. ESR-dating of calcareous deposits under- and overlying the hominid-bearing bed allowed its bracketing between 242–313 ka as upper and 147 ka as lower age limits, respectively. This result is in good agreement with uranium series data of 315–220 ka (Hennig and Grün 1983). ESR-dating was also tried on travertine, e.g., for the profiles Weimar–Ehringsdorf and Bilzingsleben, central Germany, with important hominid fossils (Schwarcz et al. 1988). The data support Middle Pleistocene ages for this site but probably need some revision using the more advanced approach now available.

ESR may also be applied to quartz grains from clastic sediments provided the grains were exposed to light for about 6 months during sedimentary transportation so that the minimum signal level in Al-centers is reached. However, there are so far only a few examples of successful applications, one being the Early Pleistocene Monte Poggio site, Italy, where more than 4,000 Paleolithic flint artifacts were found in sandy beach deposits. Detrital quartz extracted from different archeological levels provided a mean age of 1065 ± 165 ka, which is in agreement with paleomagnetic data (Falguères 2003).

10.7 Radiocarbon (\(^{14}\)C)

Among the physical dating methods, radiocarbon is the most widely known and most commonly applied one in archeology. Essentially it utilizes organic remains from the last 50 ka. In the context of paleoanthropology, this means that only the periods from the Late Middle Paleolithic onward may be dated, which comprise a relatively short section of human presence on earth. Furthermore, its application in the pre-Holocene time range (>12 ka) is complicated by the fact that reliable calibration procedures for \(^{14}\)C dates are not yet available resulting in low accuracy of the true age. Toward the upper age limit, contamination by modern radiocarbon becomes a dominating problem.

Recent natural carbon consists of three isotopes, these are the two stable isotopes \(^{12}\)C (98.89%) and \(^{13}\)C (1.11%), and in very minor traces the radioactive \(^{14}\)C (\(^{14}\)C/\(^{12}\)C ≈ 10\(^{−12}\)), which decays (\(t_{1/2} = 5,730\) a) under β-emission. It is naturally produced in the stratosphere by interaction of neutrons of the cosmic rays with atmospheric nitrogen atoms according to the reaction \(^{14}\)N(n,p)\(^{14}\)C. The average global \(^{14}\)C production rate is 7.5 kg/a but is subjected to considerable temporal variations. Depending on this rate, an equilibrium value between production and radioactive decay is established. Bound in CO\(_2\), \(^{14}\)C becomes quickly distributed throughout the atmosphere. By photosynthesis, CO\(_2\) enters the plants and via the food chain the biosphere. Through dissolution and gas
exchange it moves into the hydrosphere, from where it gets as CaCO$_3$ into marine or limnic sediments and organisms. As long as these reservoirs participate in the carbon cycle, their $^{14}$C concentration maintains the atmospheric equilibrium value. With death of the organism or with carbonate precipitation, the carbon becomes separated from the cycle so that the $^{14}$C input is interrupted. Thereupon the $^{14}$C/$^{12}$C declines due to radioactive decay. Initially it was assumed by Libby (1952) that the atmospheric $^{14}$C/$^{12}$C was constant so that according to this model the conventional $^{14}$C age is calculated (cf. Eq. 5) from the present $^{14}$C/$^{12}$C and an assumed fixed value for $(^{14}$C/$^{12}$C)$_0$. The conventional $^{14}$C age is linked to the reference year 1950 AD, with the notation “BP” (before present) after the age value.

Conventional $^{14}$C ages are incorrect since the simple model condition of a constant initial $^{14}$C/$^{12}$C is invalid. Tree-ring data, for instance, show that within the past 11 ka the initial $^{14}$C/$^{12}$C decreased by ca. 10%, so that a 10-ka old sample results in a $^{14}$C age being ca. 1 ka too young. In order to obtain accurate ages, the temporal fluctuation of $^{14}$C must be recognized. This is achieved by $^{14}$C dating of samples of independently known age. Apart from possible oscillations of the primary cosmic ray flux, the fluctuations in the atmospheric $^{14}$C/$^{12}$C level are attributed to changes in the $^{14}$C production rate caused by the varying magnetic shielding of the primary cosmic rays; in particular the geomagnetic minimum ca. 40 ka doubled the production rate (Beck et al. 2001). Also the magnetic disturbance, originating from the solar activity, modifies the production rate, although to a lesser degree and on shorter timescales. In this way, high magnetic intensities imply low $^{14}$C production rates and, thus, high apparent $^{14}$C ages. On the other hand, the carbon exchange between different reservoirs, mainly the ocean and the atmosphere, affects the atmospheric $^{14}$C/$^{12}$C level. In periods, when the oceanic circulation slows down, less of the “aged” oceanic CO$_2$ is released into the atmosphere so that atmospheric $^{14}$C/$^{12}$C level increases, leading to apparently younger $^{14}$C ages.

Regardless of their origin, these fluctuations necessitate calibration of the $^{14}$C ages. By calibration, the chronologically irrelevant conventional age is converted into a calendar age. This is achieved by calibration curves in which conventional $^{14}$C ages are plotted versus the true calendar ages. For the Holocene, this is done with high accuracy by dendro-calibration. For the Late Pleistocene, other archives than tree-rings are needed due to the scarcity of trees during glacial climate. Annually varved sediments and shallow-water corals, which can be independently dated with uranium series techniques, were used for this purpose. The results of such efforts were presented as INTCAL98 curve (Stuiver et al. 1998), which reaches back to 24,000 calendar years, and the subsequent IntCal04...
(Reimer et al. 2002). In order to extend the calibration back to 50 ka, several attempts were undertaken, using lake sediments, corals from uplifted marine terraces, submerged speleothems, and deep-sea sediments (cf. Bard et al. 2004). In this early period, the conventional $^{14}$C ages are consistently younger than the true ages by 4–5 ka. A particular complication is posed by the strong $^{14}$C fluctuations that are associated with the magnetic minima during the Laschamp and Mono Lake geomagnetic excursions around 40 and 33 ka, respectively, which actually prevent reliable $^{14}$C dating for that period (Conard and Bolus 2003). In order to distinguish calibrated $^{14}$C ages from conventional ones, they are characterized by the notations “cal BC” (calendar years BC), “cal AD” (calendar years AD) or “cal BP” (calendar years before 1950 AD).

Apart from the temporal fluctuations, there are spatial $^{14}$C nonhomogeneities in the materials and reservoirs participating in the carbon cycle. One distinguishes isotope fractionation and reservoir effects. Photosynthesis, for instance, enriches the light $^{12}$C over the heavy $^{13}$C, and in turn the latter over the even heavier $^{14}$C, so that the carbon in plants is isotopically lighter than in the atmosphere. For most materials, the age corrections that result from isotope fractionation are less than 80 a but may amount up to several hundred years, as in the case of marine limestones and organisms. The reservoir effect deals with the isotopic variation of carbon within the reservoir from which the organisms extract their carbon. Such spatial changes may have various causes. If the carbon stays long—with respect to the $^{14}$C half-life—within the same reservoir, the $^{14}$C concentration declines (“aging” of carbon). A prominent example is the “marine reservoir effect” in the oceans where upwelling regions have apparent $^{14}$C ages around 400 a. Also the admixture of “aged” carbon lowers the $^{14}$C concentration, such as “hard-water effect” in carbonaceous ground and surface waters. Reservoir effects result in an apparent increase of the $^{14}$C age. They are difficult to assess.

As for the other radiometric dating systems, the $^{14}$C system also has to remain closed, i.e., carbon must neither enter nor leave the sample. The $^{14}$C age is lowered by uptake of recent carbon. Common sources of contamination with recent carbon are the presence of rootlets, humic acid infiltration, and bioturbation. The danger of contamination by modern carbon is the greater, the smaller is the authigenic $^{14}$C amount and the older the sample. For this reason, the applicability of $^{14}$C dating at high ages is limited by unavoidable contamination rather than by the instrumental capabilities of $^{14}$C detection. The upper dating limit (maximum age) is reached when the $^{14}$C of an old sample cannot be discriminated with sufficient statistical confidence from the background. The AMS technique has the great potential to lower the instrumental background and
thus to extend the maximum age but effectively is limited by the $^{14}$C background due to contamination.

The amount of sample required for $^{14}$C dating depends on the carbon content, the conditions of preservation, the degree of contamination, and the technique of $^{14}$C detection. For $\beta$-counting, either in gas or liquid scintillation counters, 5–10 g of extracted carbon usually is needed. The AMS technique requires carbon in the milligram range. Note that the quoted amounts refer to carbon and not sample. The required amount of the latter one might be larger by a factor of 10 or so, depending on the carbon content.

**Bones** and [antler](https://en.wikipedia.org/wiki/Antler) are the most frequently used paleoanthropologic sample materials for $^{14}$C dating. As long as their inorganic fraction was used, bones were considered as a problematic material for $^{14}$C dating due to open system behavior. However, their organic substance consisting predominantly of various proteins, generally classed as collagen, is more resistant to exchange. The collagen is chemically extracted as acid-insoluble residue and is then usually subjected to AMS analysis. In an extensive program, Conard and Bolus (2003) dated numerous animal bones, the majority of them with anthropogenic modifications, from several Upper Paleolithic sites in the Swabian Jura, Germany. Among these sites were the famous Vogelherd and Geissenklösterle caves with rich finds of small figurines and flutes made of bone. The conventional $^{14}$C ages indicate that in this area, the Aurignacian spans the period between 40 and 30 ka BP and the Gravettian was well-established not later than 29 ka BP. This chronology—which actually is supported by luminescence dating of burnt flint (Richter et al. 2000)—seems to be significantly higher than those at other European sites with comparable Upper Paleolithic assemblages. However, the true chronostratigraphic position and duration of the Aurignacian cultural group cannot be directly deduced from these data due to the strong global fluctuation of the atmospheric $^{14}$C level around this period. Unfortunately, this “Middle Paleolithic Dating Anomaly” coincides roughly with the period of the Upper/Middle Paleolithic transition and the arrival of modern humans as well as the extinction of the Neanderthals, and thus seriously hampers the chronological solution of the basic questions connected with these culture and population changes (Conard and Bolus 2003). Commonly it is believed that the beginning of the Aurignacian is associated with the arrival of modern humans and among the best evidence appeared to be the presence of modern human skeletal remains within the Aurignacian layers at the Vogelherd cave. $^{14}$C dating of these bones, however, revealed that they are Holocene intrusive burials within the Paleolithic levels, leaving open whether modern humans indeed produced the Aurignacian artifacts (Conard et al. 2004).
Another important dating material is charcoal, as for instance at El Castillo cave, Spain. Charcoal samples were taken from different parts of the lowermost beds containing Aurignacian artifacts. The mean $^{14}$C-AMS age of the three samples amounts to $38.7 \pm 1.9$ ka BP (Valdes and Bischoff 1989).

Limnic sediments form an important archive for the climatic fluctuations of the past and if they contain organic matter $^{14}$C dating can be directly applied, in particular to peat and sapropel. An excellent material for AMS $^{14}$C dating are macrofossils, such as nutlets, fruit scales, and leaves, from the sediments. Such fossils are short-lived and free of the hard-water effect. When dating secondary calcareous sinter, the uptake of a certain fraction of “dead” carbon from geologically old limestone must be taken into account, which lowers the reliability of such dates. When using mollusc shells, the marine or hard-water reservoir effect, depending on the habitat of the molluscs, need to be considered. Paleolithic rock paintings commonly contain organic material, such as charcoal, carbonized plant matter, pigments, plant fibers, blood, fatty acids, and beeswax, which enables $^{14}$C dating. The recently discovered rock paintings at Chauvet-Pont d’Arc, France, gave $^{14}$C ages $\approx 31,000$ a BP, using microsamples of charcoal from the paintings (Clottes et al. 1995), and thus belong the earliest examples of prehistoric rock art so far discovered.

10.8 Conclusions

Dating methods based on radioactive decay are in principle independent clocks, i.e., they solely rely on the measurement of radiometric quantities, on known physical constants and natural isotopic abundances. The potassium–argon, fission track, luminescence, ESR, and uranium series methods are such independent clocks. Radiocarbon, on the other hand, is a dependent clock since it requires calibration by independently derived ages, for instance by dendrochronology or uranium series dating.

Radiometric dating provides the base line for calibrating other dating approaches such as isotope-, climate-, bio-, and magnetostratigraphy as well as astronomical and chemical means. All these latter techniques rely on natural changes with varying rates that are more or less predictable. Only through calibration do these techniques become chronometric tools.

During the last decades, a solid chronology for the time period hominids entered the scene since some 5 Ma ago, has been established through radiometric dating. Present methodologic developments focus primarily on improving the time resolution, which is a prerequisite in order to decipher natural and cultural
processes. Paleoanthropologic knowledge and concepts have greatly benefited from these advances.

References

Edwards RL, Chen JH, Wasserburg GJ (1986/87) $^{238}$U–$^{234}$U–$^{230}$Th–$^{232}$Th systematics and the precise measurement of time over the past 500,000 years. Earth Planet Sci Lett 81: 175–192


Richter D, Waiblinger J, Rink WJ, Wagner GA (2000) Thermoluminescence, electron spin resonance and 14C dating of the Late Middle and Early Upper paleolithic site of...


11 Geological Background of Early Hominid Sites in Africa

Ottmar Kullmer

Abstract

Hominid remains are rare elements in the fossil record. Probably the small population sizes of early hominids, in combination with preservation constraints, limit the probability of a higher frequency of fossil discoveries. The early evolution of mankind appears to be a Pan-African story, even though the distribution pattern of remains concentrates on eastern and southern Africa. More recent findings from the Chad Basin in central Africa demonstrate that fossil hominid remains are not restricted to the eastern part of Africa. The success of exploration in paleoanthropology depends on the discovery of appropriate sediment layers, mainly lake and river deposits with an upper Miocene through Pleistocene age. Knowledge of the geological framework is of great importance in evaluating the potential for fossil preservation in a certain area. To date, three major types of geological megastructures have yielded almost all fossil remains of early hominids: the East African Rift Valley (EARV), the intracratonic basin of Lake Chad, and the fossil-rich cave deposits in South Africa. Each of these regions provides a unique sedimentary setting bearing fossil-rich layers comprising a specific time-span within the last few million years.

11.1 Introduction

The fossil record of man is scanty and discoveries are rare. It seems that the farther back in time we go, the fewer the fossil remains we can expect (Cooke 1983). There are probably several factors involved in diminishing the hard evidence of our own evolution. When a mammal dies in the open, scavengers will immediately work on the carcass, scatter pieces, and probably destroy bones. The surviving parts may be exposed to the hot sun during the daytime and cooled down at night. Extreme temperature differences between day and night lead to cracking due to loss of water and reduction of organic material in the bone fibers.
Usually this process is accompanied by the activity of microorganisms and maybe also by physical transport in riverbeds or on floodplains during torrential rain falls. All of this modifies the original surface of bone or even breaks elements apart. Most of the time, the history of an animal body ends in its complete destruction, unless the bones are covered and protected, e.g., by rapid burial with sediments (Cooke 1983). Sometimes animals are trapped in cavities, like fissures or pits, and their bones are preserved articulated. But even if we have a favorable environment for rapid covering, most of the time postburial chemical and physical processes are destructive, with some rare exceptions favored by special geological environments.

For this reason, the search for hominid fossils depends at a larger scale on the geological setting because the sedimentary environment is a limiting factor upon chances of finding early hominids. Due to the multifarious influences for destruction, each fossil we collect seems to be a stroke of luck. In Africa, the hunt for the discovery of the origin of mankind concentrates on three major types of geological settings with a great potential for new discoveries. These are graben structures, cave deposits, and intracratonic basins that developed during the last 8 million years (Ma), after the hominids diverged from their last common ancestor with the African great apes. Consequently, the East African Great Rift Valley (EARV), with its widespread outcrops of fossil-rich Miocene to Pleistocene deposits, presents a unique situation for paleoanthropological field research. The Rift provided permanent water sources for our oldest ancestors and a rich fauna. A very similar situation is found in intracratonic basins like the Chad depression, occurring as relatively static systems with large lake transgression and regression cycles, varying in time, and sometimes producing fossiliferous deposits of enormous lateral extension. In contrast, the famous South African cave sites, Makapansgat, Sterkfontein, Swartkrans, and Kromdraai, were formed by extensive freshwater karstification and sedimentation processes and their complex cavities were used as shelters for hominids and probably functioned also as traps and scavenger caves.

Geological and paleoanthropological research at African hominid sites has produced a rich database during the last century. Today the geological background information influences the search for the origin of humankind, and the growing knowledge about geology helps us to understand the processes of fossilization and supports the interpretation of individual situations at fossil localities. And the application of digital geographical information systems (GIS), in combination with high-resolution satellite images, has provided a huge additional database for field investigations. Potential fossil localities, in Rift valleys, intracratonic basins, and karstified limestone, can be localized by satellite image analysis. Nevertheless, extensive surveys by experienced field teams
will continue to be the fundamental basis for the discovery of new fossil hominid remains.

### 11.2 Geological setting

The geological setting at hominid localities is not only an important source of information about environmental settings and changes through time at a specific locality or an area but is also an indicator of the probability of finding fossilized bones and is therefore an essential aspect influencing survey strategies. Moreover, it is an important aspect of the transfer of knowledge relevant to a broad understanding of human evolution as a Pan-African story.

Sedimentary formations with a high potential for fossilization are the required geological setting. These consist of deposits that are products of weathering and transport, erosion, and deposition. Sediments are a mixture of different minerals and organic materials deposited on the earth’s surface, or in caves, in interaction with the atmosphere, hydrosphere, and biosphere. Common examples are clays, silts, sands, gravels, and breccias. Due to subsidence of some hundred meters, up to a few kilometers, deposits are typically transformed into sedimentary rock, like sandstone, claystone, or limestone.

Accumulation areas of sediments are divided into marine and continental depositional space.

Even if some of the fossil sites, like Saldanha or Klasies River Mouth in South Africa, are located closely to the ocean shoreline, the most important hominid fossil sites are located in continental deposits. Potential continental bone accumulation localities are fissures, caves, swamps, tar pits, river channels, flood plains, lagoons, and lakes. Such depositional areas can be described as geomorphological units characterized by climate, size, and shape of a sedimentary basin. Geometry and composition of sediments, and the relationship between units, provides information about the milieu of deposition. Physical, chemical, and biological parameters (Table 11.1) inform us about the depositional environment and climatic conditions. These parameters have to be recorded during fieldwork. The sum of all primary characteristics of a sedimentary unit defines the sedimentary facies (Reineck and Singh 1980). The analysis of the facies permits us to understand the sedimentary factors responsible for the appearance of the deposit and the fossils.

Medium, current, wave intensity and velocity, as well as water depth, are relevant physical factors (Table 11.2); these are the important hydrodynamic conditions for the transport of animal remains. Mineral and groundwater composition, including climate, are understood to be the chemical factors of sedimentation
and fossilization. Fortunately, sediments sometimes contain biological factors, faunal and floral remains, in different stages of preservation. The appearance of animal remains in a sedimentary matrix depends on its so-called taphonomic history. The taphonomy (Efremov 1940) includes the two processes, biostratinomy and diagenesis, that describe the pathway from a carcass to a fossil. Biostratinomy (Weigelt 1927) deals with everything that happens to a carcass after an animal’s death, before the remains are buried. Afterward, the fossilization process, including mineral exchange and compaction, is called diagenesis. The fossilization process reduces information about an animal, but it also tells us a story that contains information about the paleoenvironment. The possibility that parts of a given animal will survive for some million years is relatively low due to taphonomy. Fossils are not randomly distributed. There are only a few spots on earth where we can find mammal remains, and there are even fewer places where we can discover hominid fossils. In most assemblages of large mammals from a single locality, the abundance of hominid specimens is less than 0.5%.

<table>
<thead>
<tr>
<th>Table 11.1</th>
<th>Sedimentary parameters of depositional environments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical</td>
<td>Surface features (ripple marks, load casts, shrinking cracks etc.)</td>
</tr>
<tr>
<td></td>
<td>Depositional features (horizontal/cross bedding, grain-size etc.)</td>
</tr>
<tr>
<td>Chemical</td>
<td>Crystallization (carbonates, salts, etc.)</td>
</tr>
<tr>
<td>Biological</td>
<td>Biogenetic remains (bones, teeth, shells, phytoliths, etc.)</td>
</tr>
<tr>
<td></td>
<td>Bioturbation</td>
</tr>
<tr>
<td></td>
<td>Excrements (coproliths, pellets, etc.)</td>
</tr>
<tr>
<td></td>
<td>Organic remains (plants, bacteria, etc.)</td>
</tr>
<tr>
<td></td>
<td>Bioerosion (gnawing marks, burrows, etc.)</td>
</tr>
<tr>
<td></td>
<td>Biostratification (stromatololiths, etc.)</td>
</tr>
</tbody>
</table>

After Reineck and Singh 1980.

<table>
<thead>
<tr>
<th>Table 11.2</th>
<th>Sedimentary factors of depositional environments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical</td>
<td>Medium</td>
</tr>
<tr>
<td></td>
<td>Current</td>
</tr>
<tr>
<td></td>
<td>Wave-intensity</td>
</tr>
<tr>
<td></td>
<td>Velocity</td>
</tr>
<tr>
<td></td>
<td>Water depth</td>
</tr>
<tr>
<td>Chemical</td>
<td>Mineral composition</td>
</tr>
<tr>
<td></td>
<td>Groundwater composition</td>
</tr>
<tr>
<td></td>
<td>Climate</td>
</tr>
<tr>
<td>Biological</td>
<td>Faunal remains</td>
</tr>
<tr>
<td></td>
<td>Floral remains</td>
</tr>
</tbody>
</table>

After Reineck and Singh 1980.
Nevertheless, the number of collected fossils has increased rapidly during the last decades. More and more research teams, with sophisticated survey strategies, explore for and investigate new fossil sites. Extensive sedimentologic and taphonomic analysis at known localities leads to a broad understanding of fossil site creation.

Commonly, the geographical distribution of hominid localities is used to interpret the origin of mankind and the migration patterns of early hominids (Coppens 1994). One has to consider, though, that the interpretation of the biogeography of a species or even of an evolutionary lineage, can be heavily biased by the factors mentioned earlier, as well as by others, such as those controlling the burrowing and preservation of animal remains.

White (1988) pointed out that the fossil record in eastern Africa is biased toward a representation of the watered axial basins. Classic hinterland sites lacking permanent water sources, e.g., Laetoli, are rare indeed. But this does not imply that these were minor constituents; they may just have not been discovered yet. The preservation of fossils is heavily dependent on the sedimentological environment. Fluviatile milieus do not necessarily grant a permanent burial; many of the fossil localities in eastern Africa are associated with rivers, and a migrating channel bed can alter even a lake margin or floodplain within a short time period.

Chances of finding an original bone association decrease with the transport distance from the point of origin. When discovered, the bone assemblage is likely to represent different individuals from different localities and different time periods (Hanson 1980). Time averaging, the mixing of noncontemporaneous populations, has to be considered when interpreting the history of bone assemblages. Nevertheless, beside sedimentological factors, the effects of time averaging are also dependent on stable or highly fluctuating populations (Behrensmeyer and Chapman 1993).

### 11.3 East African Rift Valley

If we look at the geographical distribution of early hominids in Africa, we recognize that many of the famous sites are aligned in a chain along eastern Africa from Ethiopia via Kenya and Tanzania, to Malawi in the southeast. Localities like Hadar, Middle Awash, Omo, Lake Turkana, Olduvai, Laetoli, and others have yielded a great number of hominid specimens during the last 50 years. These sites are positioned in a geological megastructure, the EARV (Figure 11.1). The EARV extends about 6,400 km from northeast to southeast.

The suture of the EARV follows a fault zone in the Continental crust. Plate tectonics explains the process of break-apart and rifting. The Nubian and
Somalian plates diverge along the main rift axes and result in a great graben with an average width of 30–40 km. The flanks are uplifted, while subsidence forms the graben floor. Active east–west extension leads to an enlargement of the accumulation space. A sedimentary fault basin develops as soon as the subsidence is great enough that an initial drainage system grows along the main faults. Minor rivers and alluvial fans transport material into the basin from the rift shoulders. A complex erosion–sedimentation interaction system starts in the active rift zone.

**Figure 11.1**
The geological megagraben of the EARV (gray area) with its famous hominid sites, extends over more than 6,400 km from the Gulf of Aden and the Red Sea in the north into the Indian Ocean in the southeast at the height of Madagascar. The main rift splits into eastern and western branches just north of Lake Victoria. A schematic transverse section (below) through the Rift Valley shows that the graben shoulders are composed of a stepped block-system separated by normal faults. The center is lowered due to crustal extension and divergence and uplift of the shoulders.
The expression “rift,” or “rift valley,” can be traced back to the definition of Gregory (1896). Accordingly, a rift valley, like the Malawi Rift, is a parallel-sided down-faulted valley some tens of kilometers in width and at least a few hundred kilometers in length (Ring and Betzler 1995). Several smaller-scaled tectonic structures, like half-grabens, horsts, warped blocks, major faults, transform faults, and pull-apart basins can occur within a rift system.

All these geological structures provide space for accumulation, and erosional processes to bury and unearth skeletal remains.

In the history of a mature rift system, diverse environmental settings can be recorded, representing a wide range of habitats. Rifting may end in a marine transgression phase and the birth of an ocean, when the continental breakup continues and rift floor subsidence extends toward the coast, as we can observe happening in the Red Sea and the Gulf of Aden. In other cases, the rifting dynamics may stop after some active phases.

The EARV is probably one of the most attractive rift systems, reflecting several stages of rift development in its longitudinal extension. It therefore provides an important field of research for understanding crust breakup in a continental setting. Over the last 110 years, extensive research has been focused on the EARV (Gregory 1896; Wegener 1912; Krenkel 1922), making the East African rift system one of the best documented continental rift systems on Earth (Ring and Betzler 1995). Two areas have been of major interest: the Ethiopian Afar Triangle (Baker et al. 1972; Mohr 1987) in the Northeast at the horn of Africa, because of its special rifting situation as a triple junction of three crustal plates; and the Kenyan or Gregory Rift (Baker et al. 1972, 1988; King 1978; Crossley 1979; Strecker et al. 1990), as the eastern branch of the graben system.

Extensive magmatism and volcanism accompanies sedimentation in the graben. The oldest rifting phases are recognized in the northeastern parts at the western boundary of the Sinai Peninsula marking the boundary of the Arabian plate. The Red Sea and the Gulf of Aden are probably the oldest segments of the rift system, starting with its initial uplifts and doming of the crust, probably in the Oligocene. Initial volcanism is evident as early as the Eocene (e.g., Trap series), producing widespread lava. In the Miocene, thick flood basalts produced by large shield volcanoes filled the proto-rift in the Afar region. A highly active rifting phase followed in the Miocene and produced accumulation space for extensive sedimentation of the erosional material from the rift shoulders and the rift volcanism. Sedimentation was always interrupted or accompanied by flood basalts and other volcanic eruptions, changing the landscape and sedimentation environments within the basins.

There is no doubt that due to the spreading process, the EARV witnessed some habitat changes in eastern Africa. Faulting, uplifts, and subsidence support
the development of all common fluvial processes, with meandering and braided rivers, flood plains, alluvial fans, oxbow lakes, and natural levees. Major rift lakes, like Lake Turkana, developed permanent water sources during long-term transgression phases. Regression/transgression cycles are evident through changes in sedimentary formations.

11.3.1 Afar depression

The Afar Depression in Ethiopia (Figure 11.2) is one of the world’s hotspots in two ways. The temperature can climb up to more than 45°C during daytime in the endless plains of the Afar, and fieldwork is a great challenge in the extreme African sun. Still, for paleoanthropologists and paleontologists who deal with the evolution of early humans, this investigation can be of great interest since the area of about 150,000 km² bears some of the richest hominid sites in the world. In that

Figure 11.2
The outlines of the Afar depression in the northeast of Ethiopia draw a triangle of about 150,000 km². This huge basin contains fossiliferous sediments ranging in age from the Upper Miocene to recent times. The most famous hominid localities in the Afar triangle are Hadar and the Middle Awash area. The Omo Valley is located far south in the main rift, where the Omo river crosses the border into Kenya. The fossil-rich area of Koobi Fora belongs to the Lake Turkana localities in Kenya, where thick sediment deposits surround the recent Lake. Olduvai and Laetoli are positioned within the eastern branch of the Earl, while the Chad Basin reflects an intracratonic depression about 2,500 km west of the Great Rift Valley with widespread deposits of the ancestral Lake Chad.
context, the Afar Depression contains fossils from the earliest time interval of humankind, 5–6 Ma, right up to the youngest.

Since the beginning of the 1970s, the Afar Triangle has been in the spotlight of international paleoanthropological field research. Numerous hominid findings, like the skeleton of the famous “Lucy,” come from here. Paleontological work in Ethiopia started in 1902, when the Frenchman Robert de Bourg de Bozas discovered fossil-rich deposits in the lower Omo Valley north of the Kenyan boundary (Figure 11.2).

Nevertheless, the first fossil hominid from Ethiopia, a lower jaw, was discovered only in 1933, in a cave close to Dire Dawa in the east. Likewise in 1933, the French paleontologist Camille Arambourg led an expedition to the fossiliferous sediments of the Omo river (Arambourg 1933). Although no hominid remains were discovered during the fieldtrip, the team collected numerous animal fossils (Arambourg 1947). In 1967, an Ethiopian–American–French–Kenyan Team under the direction of Clark Howell and Yves Coppens, and with the collaboration of Camille Arambourg, started the intensive search for hominid fossils in the Omo region (Arambourg et al. 1967; Arambourg and Coppens 1968a, b; Coppens et al. 1976). Up to 1973, they recovered several hominid remains in the Usno, Mursi and Shungura formations in the Omo River Basin. This initiated a “hominid-rush” in Ethiopia, which continues today.

When the French geologist Maurice Taieb in 1970 announced the discovery of fossil-rich deposits at Hadar, further north in the Afar depression, the search for hominid remains shifted into the Afar region (Taieb et al. 1972, 1974). It was at Hadar that Donald Johanson and his team found the most complete A. afarensis skeleton in 1973. Inspired by a Beatles song, they named it “Lucy” (Johanson and Edey 1981). From 1975 to 1978, the American geologist Jon Kalb explored with his colleagues of the “Addis Abeba Rift Valley Research Mission in Ethiopia” (RVRME) the sediments in the “Middle Awash” region (Kalb et al. 1982a, b), where they have documented one of the best known lithological sequences in the Afar Triangle containing fossil evidence of human evolution. The sediments cover a time-span of more than 6 Myr. Intercalated volcanic ash and lava horizons permit the absolute dating of sandwiched river and lake deposits, and thereby provide a unique insight into the evolutionary history of humankind. Since the early 1990s Tim White and his “Middle Awash Research Project” have investigated successfully in these deposits. They have collected a large mammal assemblage including numerous hominid fossils, like the first remains of Ardipithecus ramidus (White et al. 1994). Since spring 2000, the international PAR-Team (Paleoanthropological Research Team), under the direction of Horst Seidler, has investigated deposits further south of the “Middle Awash” in a fossil-rich area in the Somali region, called Galili, about 100 km
north of the Awash railway station and town in the vicinity of the rift shoulder (Macchiarelli et al. 2004).

11.3.2 Central Rift

Further South in the EARV, in the Turkana Basin of northern Kenya, Richard Leakey and his team started to explore the eastern shore of Lake Turkana from 1968 through the early 1970s (Leakey and Leakey 1978), while investigations in the Omo River Valley were continuing. They surveyed the areas in the vicinity of Koobi Fora (Figure 11.2), a place where fossil bones were reported as early as 1940, and mentioned by the District Commissioner of Marsabit to Dr. L.S.B. Leakey, Richard Leakey’s father (Leakey 1978). The Koobi Fora Research Project collected fossils including many hominid remains from Plio-Pleistocene lake, river and flood plain silt, and sand deposits that are bordered by Miocene volcanic lava in the east and south, and by the recent Lake Turkana in the west.

Extensive volcanic activity accompanying the rifting process along the major axis of the rift valley forms depositional bodies with material from ash rains and lava flows. Those tuffs, basalts, and ignimbrites are important marker horizons intercalated in the sediment successions. Radiometric dating allows the determination of their absolute age and therefore also of the sandwiched fossil-bearing sediments at many hominid sites, in a region running from north of Lake Turkana in the Omo Valley, along the eastern and western shoreline, and into the southwest at Kanapoi and Lothagam. The sediment succession tells a story of transgression and regression phases during the development of the great graben. Today sediments of different age outcrop at the same level alongside them is due to small scale tectonics that produces uplift and subsidence of single blocks leading to a patchwork of sedimentary deposits. To sort the sediments into a logical lithostratigraphic column at a particular fossil locality is one of the many challenges for sedimentologists reconstructing the sedimentary settings. It took many years to reconstruct the development of the sedimentary bodies deposited during the last 6 Myr at Lake Turkana.

Further south in the EARV, the axis of the major graben seems to be nebulous. On a larger scale, the continental fracture in the earth’s crust splits in northern Tanzania into western and eastern branches, forming a large island structure between that holds an intracratonic basin, a large depression with Lake Victoria in its centre. The western branch of the EARV, with its northern deposits at Lake Albert in Uganda, has also produced a rich fauna, eroding from lake beds and river deposits from the Upper Miocene onward (Pickford et al. 1993).
So far no hominid fossils have been discovered in these sediments, whereas in the southern part of the eastern branch the fossil localities of the famous Olduvai Gorge in northern Tanzania have produced many hominid remains. In the vicinity of Olduvai, a remarkable magmatic province, with the Ngorongoro caldera, Lemagrut, Sadiman, and Oldeani volcanoes, produced immense amounts of ash during Plio-Pleistocene times, including the thick volcanics that mark the base of the Olduvai succession. The volcanic eruptions supported the creation of a fossil site that seems to be unique in the EARV. Most likely the oldest volcano, Sadiman (K–Ar age of 3.7 Ma), was the source of the volcanic sediments of the Laetoli region (Hay 1987). The distinctive geochemical composition is compatible with the Laetolil Beds, while Lemagrut and Ngorongoro produced lavas of a different petrology. The fossiliferous Laetolil Beds, close to the village Endulen, were discovered by some local Masai. Erosional processes unearthed several hominid remains, and in addition a rich fauna was discovered. The first hominid remains at Laetoli were collected by a German explorer, Kohl-Larsen in 1939, 4 years after the first visit to Laetoli of Mary Leakey. She traveled again to Laetoli in 1959 (Leakey 1987), while she and her husband Louis were excavating in the sediments of the famous Olduvai Gorge in the North. Later in the 1970s, Mary Leakey discovered one of the most remarkable pieces of evidence for the upright gait of early hominids: a footprint trail in a volcanic ash deposit. The lowermost sediments, the Laetolil Beds, were deposited on an uplifted peneplain prior to movement along the Eyasi Fault (Hay 1987), a large fault creating the steep Lake Eyasi cliff. Some modern inselbergs give evidence of the uneven surface of the peneplain, where the sedimentary deposits in the valleys mark the paleodrainage and point to east-west as the major flowing direction. Obviously, the steep Eyasi fault did not exist at the time when the water-laid deposits developed (Hay 1981), because today the river systems in the south flow more or less perpendicular to the Eyasi fault. The Laetolil Beds deposits are divided into dominantly water-laid tuffs in the south, with the addition of extensive aeolian tuffs in the north. In the upper sequence some lava flows and clay deposits occur. In recent times doming and resulting uplift led to erosion in the Laetoli region and to the exposure of the stratigraphic sequence of fossil-rich sediments.

11.4 Chad Basin

While today water erosion and fluviatile deposition during short and heavy rainy seasons is the most active land-forming process in the eastern African savannah, the Sahara desert is dominated by wind erosion and dune deposition. Brunet
et al. (1995) reported the first *Australopithecus* remains far west of the EARV from the desert of central Africa, more than 30 years after Coppens (1966) announced the first hominid discovery in Chad. Since 1994 the Mission Paléoanthropologique Franco-Tschadienne, led by Michel Brunet has explored Miocene and Pliocene deposits in the Djurab Desert of northern Chad. The deposits are located in a large intracratonic basin, which includes modern Lake Chad in the southern subbasin and the Chad lowlands in the northern subbasin (Vignaud et al. 2002). Fossiliferous sediments in the basin have been known since 1960, when Coppens reported Quaternary fossils from Koro Toro.

Recent aeolian deflation in the northern subbasin formed the Djurab Desert, with arid conditions, while the southern depression, under semiarid to wet conditions, comprises the latest lacustrine episode in the region. A “Mega Lake Chad” existed in Holocene time, but progressive desert extension of the Sahara toward the south reduced the water column of Lake Chad to its present shallow water situation. An average water depth of 2–4 m can be measured today and must be compared to a maximum depth of 180 m during ancient times. The basement of the Chad Basin is built of Precambrian rock and forms a large depression which is filled with lacustrine, fluviatile, and lake sediments. The fluviatile facies is described by flooding channels, probably deposited during torrential rainfalls because no mature river system, like meandering or braided channels, is reported (Vignaud et al. 2002).

In the northern subbasin, extensive aeolian sands contain well-sorted quartz grains, frequently formed as dunes and shifting sands, since Upper Miocene times. The fossil content, including rich fish remains, crocodiles, and semiaquatic mammals, like hippos, gives evidence of a permanent water source during the time when *Sahelanthropus tchadensis*, the oldest known hominid, lived probably in a gallery forest along the shoreline of Proto-Lake Chad. The so-called anthropotheriid unit of the *Sahelanthropus* site is interpreted to be a shallow perilacustrine environment and consists of a well-sorted and well-cemented sandstone. It yielded all the terrestrial vertebrate remains (Vignaud et al. 2002).

However, the desert already existed in the vicinity and produced sedimentary material through aeolian deflation that gradually covered the lake floor and also animal carcasses. Along the flat margin of the ancient lakes, drainage systems developed during rainy seasons. Floods reworked the aeolian deposits surrounding the lake. Temporarily thin soil layers developed on floodplains, as confirmed by root casts developed in the major fossil layer at the *Sahelanthropus* locality Toros Menalla (Vignaud et al. 2002). The sedimentary situation in the Lake Chad Basin is controlled by transgression and regression phases of the lake and the Sahara desert.
11.5 South African caves

In South Africa, the famous hominid sites of Sterkfontein, Swartkrans, Kromdraai, and Makapansgat (Figure 11.3) are formed in a very specific geological setting, compared with the eastern African hominid localities. The fossil hominid-containing sediments accumulated in caves developed in Precambrian dolomite limestones, in Plio-Pleistocene times, through extensive karstification. The cave limestones belong to the Malmani Dolomite, which is part of the Transvaal Supergroup (Eriksson et al. 1976). The age of the Malmani dolostones, deposited in the intracratonic Transvaal Basin, is considered to be between 2.5 and 2.6 billion years (Button 1973). The thickness of the deposits reaches 1,450 m, e.g., in the Sterkfontein area (Eriksson and Truswell 1974). A large sedimentary hiatus from the Precambrian to the Late Tertiary deposits is present. Abundant faulting and folding of the dolomite limestone makes determining the exact stratigraphic position of the basic cave material difficult. The occurrence of stromatoliths in the Makapan Valley led to the idea that at least some of the limestone sediments had been formed in the intertidal zone (Eriksson et al. 1976) near the shore line of the ancient sea.

The initial cave development probably followed the fracture pattern in the dolomite rock, and extensive and deep karstification took place due to groundwater level changes. A typical feature of the Transvaal karst is the occurrence of a three-dimensional hyperphreatic maze of fissure passages (Martini et al. 2003), although large caverns did develop. As a result of erosion, cave ceilings collapsed and opened larger chambers. Carbonate solutions and mineralization built up thick travertine layers on the floors. The rich travertine deposits attracted miners at the end of the nineteenth century because of their pure calcium carbonate content. On the other hand, the carbonate is responsible for the consolidation of clastic sediments that filled up caverns through openings from the land surface. Sand, chert, dolomite, quartzite materials, and also bones were washed down into the caves. Winds probably brought fine grains, following the law of gravity. Possibly several transporting agents led to the high bone accumulation at some localities in the cave systems. Some of the caverns were completely filled up with exogenic material, and their sediments concreted to cave breccias with a fine-grained matrix and carbonatic cement. Differences in color and grain size may indicate different modes and phases of sedimentation. To access the bones, which are baked in the breccias, demands sophisticated physical and chemical preparation techniques.

Gravity and water transport, together with typical processes like collapses, solution, and remineralization, are responsible for the complex geological setting
of many of the caves. The extensive lime works in the South African caves brought thousands of bone fragments to light, although blasting operations destroyed many natural features of the caves. Consequently, reconstruction of the diagenesis of the South African cave sites was a real challenge and many aspects still need

Figure 11.3
The caves at Makapansgat belong to a large Precambrian dolomite formation (above) showing profound karstification and sedimentation during Plio-Pleistocene times. Limestone cavities (below) are filled with travertine and characteristic carbonatic bone breccias. Mining for carbonates hollowed the filled cavities and brought thousands of fossils to light. The preservation of the fossils differs from sites in the EARV due to very different processes of site formation.
further investigation. The analysis of the taphonomic history of bone accumula-
tions at the South African cave sites needs detailed on-site observations and also
lab work. Bob Brain and his team investigated many years at Swartkrans, before
he came up with his last model of cave development, where he proposed nine
diagenetic steps (Brain 1993). The formation started with a probably Miocene
cavern below the level of standing water. After the opening of the cave, surface
sediment began to accumulate inside the cave. The interpretation of the fossilif-
erous ‘pink breccia’ of the Outer Cave, which was shown to be an infilled
remnant, the Hanging Remnant Unit of Member 1, proved to be especially
time consuming.

Swartkrans Members 1–3 yielded inter alia the remains of Australopithecus
robustus, which is likely to have an age of 1.8–1 Ma. Members 4–5 are Middle
Stone Age and ca. 11,000-year-old, respectively.

Extensive investigations at Swartkrans and other South African cave sites lead
to conclusive models about the development of the deposits and their fossil
content in time and space. The experience and strategies used in the past provide
the knowledge and tools to further explore the karstified limestones in South
Africa and at other localities.

11.6 Conclusions

Although it seems to be clear that the chance of finding hominid fossils is limited
by the factors mentioned earlier, there is no doubt that many fossils still await
unearthing, either by erosion or by man. The types of sediments and the
geological contexts likely to yield hominid fossils are known. We just have to
look for the right deposits. So far, early hominids have only been discovered in
eastern and southern Africa, with the exception of the Chad Basin in central
Africa. This is probably due to the auspicious geological setting and time repre-
sented in the EARV, rather than reflecting the paleobiogeographic distribution
pattern of early hominids, although the occurrence of permanent freshwater
sources, represented, e.g., by large rift lakes, probably attracted early hominids,
because they were not capable of carrying a larger amount of water from one
location to another.

No doubt the hominid richness of the Rift sediments is unique altogether.
There is still a high potential to recover more remains at long known places
because erosion is always at work and the immense deposits in the EARV are far
from fully explored. Huge areas, for instance in the Afar Triangle and in other
places, have never been surveyed for fossils. New techniques like satellite imagery
with a resolution of a few meters on the ground are of tremendous help for
exploration work. Additionally, airborne survey can provide access to remote places such as those in the Danakil depression in the northern Afar. Nevertheless, extensive studies of satellite images will almost certainly discover new fossiliferous localities in smaller-scaled graben and basin structures somewhere in central, western or southwestern Africa. It seems just a matter of time until the first early hominid is reported from the other side of the African continent.

**Acknowledgments**

I especially thank Winfried Henke, Hartmut Rothe and Ian Tattersall for giving me the opportunity to contribute to the Handbook of Paleoanthropology. Further, I would like to thank Ian Tattersall, Winfried Henke and my colleague Oliver Sandrock for their helpful comments and for reviewing a draft version of the manuscript. Many thanks to Christine Hemm for help with the figures.

**References**


Coppens Y (1960) Le Quaternaire fossilifère de Koro-Toro (Tchad): Résultats d’une première


Weigelt J (1927) Rezente Wirbeltierleichen und ihre paläobiologische Bedeutung. Max von Weg Verlag, Leipzig


12 Paleoclimate

Keith Alverson

Abstract
This chapter provides a brief overview of the dynamics and pattern of climate variability from the Plio-Pleistocene to the modern day including Quaternary ice ages, millennial scale variability, abrupt climate change, and the Holocene. Selected examples of terrestrial, marine, and cryospheric archives of paleoclimate information and the proxy climate information that can be quantitatively extracted from them are presented. Although the primary focus of the chapter is climate variability itself, a few examples of the past interactions between climate and human societies are also presented. The chapter concludes with some speculative comments on lessons that can be drawn from such studies of the past with relevance to modern day concerns.

12.1 Introduction
In 2004, three major hurricanes crossed Polk County, Florida. Local people’s lives were changed in dramatic and memorable ways. The events were also felt on a national scale, affecting a number of large institutions such as the federal government and the insurance industry. From the perspective of residents of Polk County, 2004 was a remarkable climatic year. But from the temporal perspective of centuries, or the spatial perspective of the entire planet, this series of events is an insignificant footnote.

For most of human history, individuals and societies have experienced climate on local space scales and timescales of a few generations at most. These scales of human perception of climate are very different from those inherent to the climate system itself. The climate system is globally interconnected and varies on all timescales, with the most energetic variations on timescales of 100,000 years and longer. In part as a result of this disparity of scales, understanding the influence of climate on the evolution of Homo sapiens and the development of human societies is not an easy task. However, it is unquestionable that enormous climatic changes occurred on the timescales of human evolution and societal development and, at times, played a role in these processes.
This chapter focuses primarily on past climatic variability during the Quaternary including both reconstructions of past climate variability and some hypotheses seeking to explain the dynamics of this variability. In light of the likely interests of readers of this paleoanthropological handbook, a few examples of interactions between past climate variability and human societies are highlighted. The chapter begins with a brief discussion of the climate of the Plio-Pleistocene, followed by discussion of Quaternary ice ages, millennial scale variability, abrupt climate change, and the Holocene. It concludes with some speculative comments on lessons for modern day concerns that can perhaps be drawn from investigations of past climatic variability and its influence on humans.

12.2 Pre-Quaternary climate

Pre-Quaternary climate and environment conditions are not discussed in detail in this chapter but are included briefly to provide a general context for the subsequent sections. For longer time perspectives, including the paleoclimatic context of hominid evolution, the reader is referred to the chapter 16 by Etter (Patterns of Diversification and Extinction) of this handbook. One of the best indicators for global climatic conditions on timescales of millions of years is provided by the ratio of oxygen isotopes recorded in shells of deep-sea dwelling foraminifera preserved in deep-sea sediments. Figure 12.1 shows one such record from a deep core drilled in the equatorial Pacific (Mix et al. 1995a, b). Such records are influenced by a number of local variables, including for example seawater temperature, but are nonetheless thought to provide an estimate of global atmospheric temperature and the volume of water stored in continental ice. Thus, Figure 12.1 illustrates a time history of the transition from relatively warm, conditions of the late Miocene with little or no Northern Hemisphere glaciation, into the predominantly cold and glaciated, but highly variable, conditions of the late Quaternary.

Records such as the one shown in Figure 12.1 highlight a number of intriguing questions about the nature and variability of pre-Quaternary climate dynamics. The primary driver of both the overall trend and much of the variability is thought to be the pattern of solar radiation incident on the Earth. Variations of the eccentricity of the Earth’s orbit, from more circular to more elliptical, influence the degree of seasonality of climate, with opposite forcing in the northern and southern hemispheres. Over the past 5 million years, the eccentricity in earth’s orbit varied quasiperiodically on timescales of 413,000 and 100,000 years. At the same time, changes in the tilt of the earth’s axis, or
obliquity, varied with a fairly periodic oscillation on a timescale of 41,000 years. The effect of obliquity changes is primarily on the amount of solar radiation reaching high latitudes and has the same sign in both hemispheres. Finally, precession of the equinoxes, i.e., changes in the location of the earth within its elliptical orbit relative to the seasons varies with a primary period of about 22,000 years. Individual orbital forcing terms are not independent of one another, for example, the influence of precessional changes is coupled to eccentricity variations, the location of the earth within its orbit at a certain time of year being most important at times when the orbit is most elliptical.

Although changes in orbital parameters are the main external driver of paleoclimatic changes over the past several million years, internal dynamics of the climate system itself are far more important if one is interested in the actual temporal and spatial patterns of climatic change. For example, orbital changes do not explain why the global climate system responded primarily to the relatively weak 100,000 year forcing associated with eccentricity during the Quaternary, but to the 41,000 year forcing associated with obliquity during the late Pliocene and early Pleistocene. Nor do they explain the long-term cooling trend starting in the early Pliocene. Relevant processes internal to the Earth system that influence the global climate state on these timescales include concentrations of atmospheric greenhouse gases, the pattern and degree of reflectivity of the earth associated with changes in the biosphere and cryosphere, and orographic changes associated with tectonics.
One example of climatic variability driven primarily by a combination of external and internal system dynamics is the development and variability of the East Asian Monsoon. Conveniently, one of the most well-preserved terrestrial archives of climatic change over the Plio-Pleistocene come from loess deposits in central China. These widespread 100–300-m thick deposits are characterized by alternating layers of loess dust and interbedded paleosols. The loess was deposited during relatively cool periods dominated by northerly winter winds while the paleosols reflect pedogenesis associated with relatively warm conditions and the moisture bearing summer monsoon. The time series of monsoonal fluctuations captured in loess records has been extensively correlated with deep-sea oxygen isotope records and, not surprisingly, cool dry periods of strong loess deposition tend to correlate fairly well with periods of high global ice volume. As shown in the Lingtai loess section in Figure 12.2 (An 2000), magnetostratigraphic measurements suggest an onset of loess deposition, and thus the inception of the East Asian monsoon, may have been as early as 7 Ma. The uplift of the Tibetan Plateau, a primary driver of continental aridity, may have already been substantial at that time. Loess records also indicate evidence for a later pulsed uplift in the Tibetan plateau around 3.5 million years ago, again consistent with the onset of strong northern hemisphere glaciation indicated by marine records.

Although intriguing, when thinking about paleoclimate records of the Plio-Pleistocene such as those shown in Figures 12.1 and 12.2, it is important to keep in mind several substantial caveats. Because there have been an enormous number of poorly understood processes involved in the development, preservation, and measuring of these archives, there remains enormous uncertainty in interpreting deceptively precise measurements of material properties such as oxygen isotopes and magnetic susceptibility in terms of anything beyond a qualitative impression of climatic parameters such as global temperature, ice volume, or East Asian monsoon strength. Quantitative estimates, or uncertainty ranges, provided in units such as temperature in degrees centigrade or ice volume in cubic meters, are usually, for good reason, avoided altogether. A relative paucity of climate proxies on these timescales makes it difficult to independently verify many of our interpretations of those we do have without employing circular logic. For example, the records are not absolutely dated, other than perhaps at a few critical points, and are usually tuned either to other records or orbital parameters, leading to obvious logical circularity. Although these types of problems are common to all paleoclimatic reconstructions, on all timescales, they become substantially less severe for Quaternary reconstructions due, largely, to the availability of ice cores.
Figure 12.2
Magnetostratigraphy, lithology, magnetic susceptibility, and grain size variations of the Lingtai loess section (35°04′N, 107°39′E) covering the past several million years. A shift to stronger monsoon occurred approximately 3.5 million years ago. Marked variability on 41,000-year timescales in the early Pleistocene followed by dramatically dominant 100,000 cycles of the Quaternary ice ages is also evident (An 2000).
12.3 Quaternary glacial–interglacial cycles

The climate of the last several hundred thousand years has been dominated by enormous glacial–interglacial oscillations on a period of about 100,000 years. During glacial maxima, vast expanses of the Northern extratropics, essentially all of modern day Canada for example, were buried under ice, whereas during relatively brief interglacial periods high latitude northern hemisphere was warmer, with glaciers restricted to high altitudes and extreme polar environments. From a paleoclimatological perspective, we are rather lucky that interglacials have not been a bit warmer and thus still supported year round ice in polar areas, notably Antarctica and Greenland. It is deep cores from these ice masses that provide continuous, quantitative records of temperature and atmospheric properties. One unique property of ice cores as a tool for paleoclimatic reconstruction is that, in addition to standard “proxy” measures of paleoclimate, such as inferring past temperature from oxygen isotope measurements, ice cores also trap actual atmospheric samples from past times in bubbles in the ice. Measurements of properties of these tiny gas samples, such as carbon dioxide concentration, provide extremely powerful evidence for paleoclimatic analysis, analogous to fossils in the archeological record.

The longest ice core records come from Antarctica, where as many as eight glacial–interglacial cycles are reflected over a continuous seven hundred thousand year record. Deuterium, an isotopic proxy for temperature, as measured in one such core is shown in Figure 12.3 (EPICA Community Members, 2004). At the writing of this chapter, a carbon dioxide record over the full length of this record had not yet been published.

Over the past few glacial cycles, a wealth of paleoclimate evidence from both hemispheres provides a coherent picture of massive 100 thousand year glacial–interglacial oscillations. On these timescales, high latitude northern hemisphere summer insolation is the most plausible primary driver of global atmospheric temperature oscillations, depending on location and with substantial uncertainties depending on the method of reconstruction, in the range of 2°C–20°C. Dramatic well-documented associated changes occurred in the terrestrial biosphere and landcover, affecting for example surface reflectivity and global water pathways; in ocean surface temperatures, affecting greenhouse gas uptake; and in atmospheric dust and greenhouse gas concentrations, affecting radiation balance. Such changes are among many strong feedback mechanisms that potentially contribute to explaining why relatively minor radiation receipts over a relatively miniscule area should have resulted in a substantial response over the entire globe.

A compendium of a few sample proxy records providing a flavor for nature and magnitude of climatic and environmental on glacial–interglacial timescales.
in both hemispheres is presented in Figure 12.4 (Labeyrie et al. 2003). From an anthropological perspective, some of the most intriguing changes include oscillations in global sea level associated with trapping of water in continental glaciers and changes in terrestrial ecosystems associated with changes in temperature and moisture availability. For example, the sea level decrease of well over 120 m at the last glacial maximum lead to formation of a land bridge facilitating relatively easy human migration from Asia to the Americas and climate change must have played a role, alongside contemporaneous hunting pressures from humans, in the well-established subsequent decline of North American megafauna.

One of the most intriguing dynamical questions associated with Quaternary ice ages is why atmospheric CO₂ varied in such tight correlation with temperature. This correlation presents an intriguing set of questions about the operation of climate, the Earth system more broadly and in particular the global carbon cycle.

Some studies seeking to address these questions have focused on what sets the upper and lower limiting values of atmospheric CO₂ concentration of roughly 180 and 300 ppm during glacial and interglacials, respectively (Falkowski et al. 2000). Slowdown of the terrestrial ecosystem during glacial periods leading to reduced ability of the terrestrial biosphere to take up carbon has been implicated as a possible explanation for the lower bound, while the upper bound might simply be a response to global cooling associated with glacial inception after relatively brief interglacial periods, with oceanic processes setting the range and frequency of variability on timescales longer than centuries.

Figure 12.3
Deuterium, a proxy for temperature, over eight glacial–interglacial cycles as measured in the EPICA Dome C ice core (EPICA 2004)
Several oceanographic mechanisms have been proposed to explain, either individually or in concert, the decrease in atmospheric CO₂ associated with glacial states as compared to interglacials. These include: increased efficiency of the “biological pump,” in which net carbon uptake, especially by the high latitude marine biosphere, is increased relative to upwelling fluxes, changes in the rate of ventilation of deep water in the Southern Ocean, decreased global thermohaline overturning circulation, enhanced air–sea exchange in the northern Atlantic, decreased air–sea exchange in the Southern ocean due to enhanced sea ice cover, increased solubility of CO₂ in colder seawater, and a middepth chemical divide separating watermasses with low and high CO₂ concentrations (Pedersen et al. 2003).

Some studies suggest simple, albeit highly specific, mechanisms (Stephens and Keeling 2000), while others claim that all simple mechanisms can be eliminated from consideration (Pedersen and Bertrand 2000), indicating that multiple or unknown processes must be at work. In contrast, LeGrand and Alverson (2001) assert that the problem is underdetermined and that many plausible solutions therefore exist. Employing a simple inverse box model, they find several solutions. Within reasonable uncertainties, a single change in the Southern Ocean relative to their modern control simulation, for example reduced air–sea gas flux, reduced deep water ventilation, or enhanced biological productivity, can explain all of the model and paleodata constraints available for the glacial state.

What can we conclude about the global carbon cycle? Clearly, the upper bound of around 300 ppm that has marked atmospheric CO₂ variability over at least the past 400,000 years is not an impenetrable one. It has already been shattered by modern anthropogenic inputs of fossil carbon to the atmosphere and is rising fast. Will this perturbation disrupt the system sufficiently to drag the Earth out of the attractor, marked by glacial oscillations, that it has been in during the Quaternary? Unless we understand how the system as it has operated in the past, it will remain impossible to move beyond simple conjecture based on correlation. Such conjecture is particularly tenuous when extrapolated to atmospheric CO₂ levels, such as those anticipated for 2,100, for which there is no analog in the Quaternary period for which well-dated, relatively precise paleorecords exist.

12.4 Millennial variability

Even a cursory glance at the records in Figure 12.4 reveals that climate variability on the timescale of millennia, particularly during glacial periods, is substantial, with larger oscillations reaching as much as half the size of the
Figure 12.4
Climatic proxies from both hemispheres over four glacial cycles as recorded in a variety of paleoclimate records. (a) Summer insolation (21 June) at 65°N. (b) North Atlantic Sea surface temperature (SST) record from deep-sea sediment core at ODP site 980. (c) South Indian ocean SST derived from foraminiferal transfer functions, from deep-sea sediment core MD 94-101. (d). Atmospheric CO₂ and temperature estimated from δD from the Vostok ice core. (e) Summer insolation (December 21) at 65°S (Labeyrie et al. 2003)
The glacial–interglacial signal itself. However, it was not until the mid-1980s that this variability was widely recognized in the paleoclimate community, initially due to the advent of high resolution ice core records from Greenland (Dansgaard et al. 1984) later corroborated by interpretation of alternating layers of “ice rafted debris” in deep-sea sediment records from the North Atlantic ocean (Heinrich 1988). In honor of these early discoverers, the half a dozen or so larger climatic swings between 60 and 15 ka are commonly referred to as “Heinrich events” while more than 20 smaller events are generally called “Dansgaard–Oeschger (D/O) events.” Moreover, the events themselves seem to be grouped in cycles, sometimes called Bond Cycles, beginning with the most severe and gradually declining in amplitude until the next multiple cycle is initiated.

By combining the stratigraphic record of these rapid D/O oscillations with attempts to model the mechanisms responsible, the dominant hypothesis to emerge sees the events as a product of the instability of the ice sheets that girdled the North Atlantic. Rapid discharge of ice into the ocean is thought to have reduced surface seawater salinity to the point where North Atlantic Deep Water (NADW) formation was inhibited, thereby shutting down the related transport of heat into the region by the large scale overturning circulation of the North Atlantic ocean. This cut-off of NADW formation may then have so modified ocean circulation as to have influenced climate on a global scale. Such a hypothesis is consistent with the antiphase relationship between Greenland and much of Antarctica often referred to as a bipolar “see-saw.”

It seems possible that the variations in amplitude of the cold “swings” reflect changes between iceberg surges reflecting instability of the Fennoscandian ice sheet, with a relatively more rapid response time, and the Laurentide Ice sheet with a longer response time. Surges of the latter are thought to have been responsible for the most extreme events, those that gave rise to the main “Heinrich” IRD layers. The periodicity of the D/O events is not regular, and no credible external forcing mechanism has been proposed. The bipolar antiphase relationship demonstrated for the major D/O events suggests that they are largely the product of cryosphere-ocean dynamics.

In the decades following the initial discoveries the same millennial events seen in the northern North Atlantic region have been found to be widespread, with clear evidence in such diverse locations as Chinese loess deposits and stalagmite isotope records, deep sea sediments from the Cariaco and Santa Barbara Basins, and even Antarctic ice cores. A compendium of a few such proxy records providing a flavor for nature and magnitude of climatic and environmental on millennial timescales in high and low latitudes and both hemispheres is presented in Figure 12.5 (Labeyrie et al. 2003). The global nature of these events has lead to further questions and refinements regarding the
possible atmospheric and oceanic mechanisms for transmitting these signals around the globe, for example, through influence on the Asian monsoon systems in the case of Chinese records, or the rate of ventilation of intermediate depth waters of the Pacific, in the case of the Santa Barbara Basin record. However, changing rates of overturning circulation driven by fresh water fluxes in the northern North Atlantic remain a widely accepted, plausible hypothesis for the primary driver of this variability.

12.5 Abrupt climate change

One of the most exciting new developments within the area of global change science has arisen from the growing body of paleoclimatological evidence for rapid changes in the climate system. Until recently it would have been hard to imagine that the global climate, together with other aspects of the Earth system to which it is intimately linked, could have undergone repeated oscillations on millennial timescales as described earlier, let alone that the most dramatic of these “switches” occurred within a matter of decades, with major consequences for temperatures, hydrological regimes, plants, and animals that were at the very least hemispheric in their extent. The evidence now available provides a fascinating insight into the climatic backdrop to the development of human societies in the past. It also highlights one of the major challenges for the future. If the potentially disastrous impacts of climatic “surprises” like the ones that have already been well documented from ice core and sediment records are to be avoided in the future, research devoted to deepening our understanding of abrupt changes in the past is of paramount importance.

Any definition of “abrupt” climate changes is necessarily somewhat subjective, as it depends in large measure on the sample interval used in a particular study and the pattern of longer-term variation within which the sudden shift is embedded. Here, I make no attempt at a general definition but focus attention on two examples of rapid transition found in the paleorecord from the geologically recent past, glacial terminations and the Younger Dryas.

12.5.1 Glacial terminations

Marine, polar ice core, and terrestrial records all highlight the sudden and dramatic nature of glacial terminations, the shifts in global climate that occurred as the world passed from dominantly glacial to interglacial conditions. The dramatic nature of these terminations is immediately evident in Figure 12.4.
One metric for the rapidity of these glacial terminations is provided by very well-documented rates of global sea level rise. During glacial terminations, sea level rose roughly 100 m in less than 10,000 years, a mean rate of roughly 1 m per century. This is at least an order of magnitude faster than anything that has been suggested for future climate change scenarios associated with global warming (Houghton 2004). In addition, during the last interglacial, sea level was about 20 m higher than today. Thus, both the magnitudes and rates of sea level rise have been significantly larger than either those that have been documented in recent years or predicted for future climate change scenarios.

Because the melting of large, northern hemisphere, continental ice sheets which no longer exist is thought to be the reason that past rates of sea level rise were so large, one might be tempted to think that understanding the controls of sea level on these timescales is not relevant to the future. There are two strong arguments against this line of reasoning. First, and most importantly, large and growing coastal populations in the modern day are highly vulnerable to sea level rise, whereas during the last interglacial no such settlements existed. Thus, even a much more gradual, smaller amplitude sea level rise occurring today will have much greater human impact than those of the past. Furthermore, extant ice sheets on Greenland and Antarctica represent fresh water storage equivalent to many meters and many tens of meters of sea level, respectively. The possibility of rapid destabilization of even a small part of these ice sheets in the future, although less likely than many other climate change predictions, cannot be excluded from consideration. Thus, an understanding of the dynamics of large ice sheets... gained through investigation of their behavior in the past, may be important for our future.

Given the growing, and increasingly urbanized, coastal population and the great potential for nonlinear “threshold crossing” response of coastal zones and the ecosystems they support to sea level rise, there is an urgent need to understand the processes driving and responding to sea level rise on timescales from decades to millennia. Many of these processes, such as postglacial rebound of continental plates, continue to effect sea level change today. The past record of sea level change is thus highly relevant to its likely future course.

Although sea level is a broadly global property, two features of glacial terminations tend to qualify the view that they are a single, rapid, globally coherent, monotonic process. When the inferred temperature change in Antarctica and Greenland are carefully synchronized on the basis of their records of changing atmospheric methane concentrations, it is clear that during the course of the warming trend from Glacial to Holocene, the record from Antarctica leads that from Greenland; moreover during the period of warming, the sharp fluctuations in temperature that mark the Greenland record in particular, are largely in
antiphase with the changes in much of Antarctica. Recent high resolution Urani-
umum Thorium dating reinforces the idea of an Antarctic lead, since it suggests
that the penultimate glacial termination occurred $135 \pm 2.5$ ka (Henderson and
Slowey 2000). This date precedes the Northern Hemisphere June insolation peak
by nearly 10,000 years. Thus, although consistent with primary deglaciation
forcing by insolation levels in the southern hemisphere, or the tropics, this
evidence runs counter to the view that insolation forcing at high northern
latitudes was the initial trigger for deglaciation. Any concept of a rapid shift
associated with glacial terminations thus has to be qualified by the realization that
the rate of change is not uniform over the whole globe and that the oscillations
that are superimposed on the trend are not globally parallel in sign.

12.5.2 Younger Dryas

The Younger Dryas marks the final stage of the last glacial termination, the
transition to the Holocene. The transition out of the Younger Dryas cold reversal
into the Holocene is one of the most well-studied rapid climate transitions, both
in terms of regional climate expression, primarily in the North Atlantic and
Europe but also elsewhere, but also its influence on ecosystems (Amman and
Oldfield 2000). Some sense of the suddenness of this transition may be obtained
from the record of ice accumulation and temperature in central Greenland
shown in Figure 12.6 (Dansgaard et al. 1993). At this location, most of the
transition, a near doubling of snow accumulation rate and a temperature shift
of around $10^\circ$C, was accomplished within a decade or less. By any standards,
this is a truly dramatic, abrupt shift in climate, yet the generally accepted external
forcing, an increase in solar irradiance, was relatively modest and certainly
gradual. Again, we see external forcing pushing the Earth system over some
crucial threshold beyond which internal dynamics can generate extremely rapid

Figure 12.5
Millennial climate variability during the Last Glacial period (60–15 ka). The timing of the
large Heinrich events is indicated by the labels H2 to H5, while the vertical gray lines mark
subsequent warming in the Northern North Atlantic. (a) Air temperature over Greenland as
inferred from ice core oxygen isotope measurements. (b) North Atlantic Sea Surface
Temperature inferred from planktonic foraminifera based transfer functions. (c) North–
South shifts in the polar front in the midlatitude Atlantic inferred from foraminifera isotope
measurements, (d,e) Deep ocean ventilation and overturning inferred from carbon and
oxygen isotope measurements in benthic foraminifera. (f) Regional temperature and mois-
ture balance changes in Eastern China inferred from oxygen isotope measurements in cave
deposits. (g) Air temperature over West Antarctica from oxygen isotope measurements in
the Byrd ice core (Labeyrie et al. 2003)
Figure 12.5 Continued

(a) Greenland ice: GISP-2 $\delta^{18}$O, 72°N
   (air temperature over Greenland)

(b) Atlantic ocean, 59°N
   - SO26-5

(c) Atlantic ocean, 37°N
   - MD95-2042
   - Planktic foraminiferal $\delta^{18}$O (N/S shifts of the Polar Front)
   - Benthic foraminiferal $\delta^{13}$C (deep water ventilation)

(d) Hulu Cave, 32°N

(e) Antarctic ice: Byrd $\delta^{18}$O, 80°S
   (air temperature over West Antarctica)

(f) Vegetation

(g) Palynology

(h) Marine sediments

(i) Oceanic circulation

(j) Paleoclimate

(k) Ice volume

(l) Sea level

(m) Sediment composition

(n) Paleomagnetic

(o) Paleoclimatic model

(p) Ice sheet dynamics

(q) Glacial history

(r) Ice core analysis

(s) Radiocarbon dating

(t) Stable isotope ratios

(u) Paleoceanography

(v) Palaeoclimatology

(w) Paleoenvironmental reconstruction

(x) Paleoclimatic indicators

(y) Paleoclimatic archives

(z) Paleoenvironmental proxies

{ka BP}
responses through a range of, as yet poorly understood or quantified, feedback mechanisms.

### 12.6 Holocene

It has sometimes been claimed that the stable climate of the Holocene helped the transition from hunter gather to agriculturally based and later development of modern societies to occur. This line of thinking arose, at least in part, due to the almost complete lack of variability in Greenland ice core isotope records during
the last 10,000 years as compared to the dramatic variability during earlier times (Figure 12.6). The paradigm can be criticized in two important ways. First, and most importantly, people do not live in central Greenland and second, there is now ample evidence from both Greenland and other parts of the world that the Holocene has not been climatically stable at all (Mackay et al. 2003; Battarbee et al. 2004). In addition, there are good reasons to believe that climatic variability, not stability, might be expected to lead to changing societal structures.

We know, from well-dated quantitative paleorecords, that regional hydrological change during the Holocene has been extremely large. The extent of glaciation in the Swiss Alps, for example, has been highly variable during this time, including periods when the glaciers were substantially more restricted than today. Lake levels in much of semiarid Africa have risen and fallen by 100 m on similar timescales. These large, and often rapid, climatic changes had substantial impacts on human societies, even playing a role in the demise of several civilizations including the Mayan, Akkadian, and Greenland Norse. Societies, develop, adapt, and change in response to variability, not when ensconced in a cocoon of stability. Indeed, precisely societal change and adaptation are the presept requirement, in the face of global anthropogenic change.

12.6.1 Early Holocene

Although in stratigraphic terms the Glacial/Holocene boundary is, over much of the world, a clearly recognizable and apparently synchronous feature, this does not imply that the Earth system as a whole experienced an instantaneous and complete shift. Several of the responses to the rapid changes taking place at the boundary took centuries, even millennia to complete. These include processes such as the melting of the polar ice caps, the recovery of global sea level to something approaching its present height, the recolonization of extensive areas by vegetation adapted to changed thermal and hydrological regimes, and the maturation of soils that goes hand-in-hand with increasingly stable vegetation cover where this had been absent during glacial times. Hence it is important to remember that the period was one of transitions and readjustments.

There is clear evidence from several quite widely spaced sites for a sudden climate anomaly lasting for several hundred years centered around 8,200 Br. This event shows up as a sharp decline in inferred temperature and in methane concentrations in ice core records from Central Greenland. An almost perfectly parallel change in d18O-derived temperature can be seen in the evidence from Ammersee, in southwest Germany, based on stable isotope analyses of ostracod
remains (Von Grafenstein et al. 1998). A growing number of records appear to indicate that this oscillation had a widespread effect. In terms of timing and signature in the Greenland ice cores records, it appears to be in many ways comparable to the preceding, higher amplitude D/O oscillations and some link to rapid draining of glacial lakes in North America and NADW formation may be involved (Barber et al. 1999), but there is also the possibility that the event, especially the methane anomaly, is linked to lower latitude changes in the extent of tropical and subtropical wetlands or the effects of rising sea level. Certainly, the whole of the early Holocene is marked by dramatic shifts in lake level and wetland extent in Africa and Central America.

The records of changing lake levels for a range of sites in Africa constitute dramatic evidence of the major changes in hydrology that occurred during the early Holocene. Two examples of such lake level records are shown in Figure 12.7 (Gasse 2000; Verschuren et al. 2000). Equally powerful indications of variability come from reconstructions of the changing extent of wetlands and vegetation cover in the Sahara/Sahel region and from evidence for high altitude lake level variations in the Altiplano of the central Andes. The dynamics of these dramatic, low-latitude hydrological changes are complicated by the fact that the changes in surface hydrology and vegetation have a strong feedback on regional climate. For example, without including this feedback alongside other forcing, including sea-surface temperatures, it is impossible for existing numerical models to simulate the extent of wetlands and plant cover in the Sahara/Sahel region during much of the early Mid-Holocene (Claussen et al. 1999).

12.6.2 Late Holocene

The later part of the Holocene is by far the most well-documented period of paleoclimatic history, this is largely due to the presence of numerous, independent, and well-preserved archives which can be precisely dated and measured with annual resolution. Notable among such extremely precise paleoclimatic records are documentary archives, rings from living and fossil trees, and growth bands from speleothems (cave deposits) and tropical corals, and annual accumulation layers from high altitude ice cores. Inevitably, attention focuses on the so-called Little Ice Age and Medieval Warm Period. Well chronologically constrained evidence for climate changes broadly paralleling, but rarely perfectly synchronous with, the European Medieval Warm Period—Little Ice Age sequence has been identified in records right across the northern hemisphere (Alverson and Solomina 2004) and into low latitudes—for example, in ice core records from Quelccaya Ice Cap in Peru.
Was the Medieval Warm Period as warm as the 1990s? Long tree ring chronologies are one of the most important sources of information available to address this question. *Figure 12.8* shows a recent, extratropical, treering based reconstruction (Esper et al. 2002) compared with an earlier hemispheric, multiproxy reconstruction (Mann et al. 1999) and their respective uncertainty.
estimates. The differences between these curves have been the subject of an active debate in the community. Of particular interest has been the question of whether the Northern Hemisphere average warming measured in recent decades is indeed greater than that associated with the peak of the Medieval Warm Period approximately 1000 years ago. The amount of cooling that occurred several centuries later during the Little Ice Age and whether this cooling was geographically restricted to certain regions in the extratropics is another issue of considerable debate.

Notable among many differences in the construction of the curves in Figure 12.8, the former (Mann et al. 1999) contains records from multiple proxies and from the tropics, whereas the later (Esper et al. 2002) makes use of a novel technique (Regional Curve Standardization) designed specifically to obviate the inherent loss of centennial scale variability in long chronologies constructed from a series of short, detrended records spliced together. Rather than highlighting the differences between these two curves, one might consider their similarities. They capture many of the same decadal scale events and lie at least 95% within each others’ 95% confidence limits. They both show the remarkable power of long proxy-climate records to put modern changes in the perspective of the past and to use this perspective to better understand natural and anthropogenic drivers of global change. Together, they also indicate some of the most promising areas for future research: better calibration of the growth response of trees to climatic forcing, the need to quantify the range of frequencies that can be reasonably expected to be captured by given reconstructions of past climate variability, and the need to develop long, annually resolved, quantitative, paleoclimatic proxy records from the tropics.
Studies of the relationship between Holocene climate variability and human societies have often been marked by antithetical perceptions within the social science and physical science communities (Oldfield and Dearing 2003). Archaeological and anthropological research encompass interpretations of socioeconomic and cultural change, resource use, and subsistence practices (Pringle 1997; Redman 1999), but the direct evidence for potentially damaging climate change is usually derived independently from different archives and lines of evidence (Cullen et al. 2000; Hodell et al. 2001). As a result, even where temporal correlations can be proposed between major societal changes and shifts in climate, they could be viewed as little more than coincidences. Taken to its extreme, the “cultural” view attributes major changes in past societies, even the collapse of ancient civilizations, entirely to human actions. Although human actions are clearly important, too one-sided interpretation is not supported by research which ascribes the collapse of civilizations as diverse as the classic Akkadian (Cullen et al. 2000), Mayan (Hodell et al. 1995, 2001), and Anasazi (Dean et al. 1985, 1999) cultures to abrupt and persistent climatic changes. Figure 12.9 presents a selection of paleoclimate examples where the combination of environmental and cultural history, coupled with rigorous chronological constraints, points to a strong link between the incidence of drought and the collapse of human cultures. Such studies do not discount the role of societal factors but assert that, at times, climate variability has been a critical factor influencing societal stability.

Four thousand years ago Yu the Great, founder of the first, semimythical, Xia dynasty of China is said to have spent more than a decade battling floods continuously inundating the valley of the Yellow river. According to Confucius’ Shu Jing (Book of History), written much later in the fifth century BC but supposedly collated from ancient sources, Yu’s great insight was to discontinue the practice of building ever-higher levees and dikes and instead dig a vast network channels to drain the great volumes of water east to the sea. Flooding, up until the present day, has continued to be a major concern for the region. A million people are thought to have died when the Yangtze broke through flood defenses in 1877. But flooding is not the only problem; the flow of the Yellow River has been in continual decline for the last 50 years. In 1997, the river ran dry in places during 226 days of the year. Much of China’s agriculture and population depends on water from the Yangtze and is greatly vulnerable to drought.

The proposed modern day solutions, multibillion dollar engineering projects, are not so very different from those proposed in the time of Yu the great.
One, the famous Three Gorges Dam, in addition to generating electricity, is intended to control flooding. A second, intended to ameliorate vulnerability to drought, is to divert water to the north, primarily from the Yangtze river basin, by digging a series of south to north canals in the western, central, and eastern parts of China. Will these gigantic engineering projects fulfill their goals? Will they reduce human vulnerability to climatic variability and change?

![Figure 12.9](image)

Two examples of paleorecords where the combination of environmental and cultural history, coupled with rigorous chronological constraints, point to a strong link between the incidence of drought and the collapse of human cultures. In the upper panel, drought in Yucatan, Mexico inferred from changes in the stable isotope ratios in two species of ostracod is shown to coincide with the collapse of Mayan civilization (Hodell et al. 2001). In the lower panel, a steep fall in carbonate percentage in a marine sediment record from the Gulf of Oman, representing a major episode of dust deposition, is directly linked to drought conditions associated with the demise of the Akkadian civilization (Cullen et al. 2000; Adapted from Alverson and Oldfield 2003)
The answers to these questions are a matter of much debate, but one salient fact is often overlooked in that debate. Paleoenvironmental science provides a detailed, quantitative record of human society, climate variability, and their interaction over the past several thousand years, both in the Yangtze valley region and over the rest of the globe. Such information from the past is directly relevant to questions of climate variability and human vulnerability today and therefore directly relevant to the modern debate.

12.7 Anthropocene: Implications for the future

Understanding past climatic and environmental change is clearly relevant to predicting the future. This contention is not new (Alverson et al. 2000, 2003; Alverson and Kull 2003), but a few salient examples deserve highlighting here.

- The paleorecord extends instrumental climate records thereby increasing the statistical significance of documented anthropogenic change.
- Long-term processes in the paleorecord operating on timescales of decades to millennia are still occurring and will influence the present and the future as they have in the past.
- The paleorecord contains a much fuller representation of the range of climatic and environmental variability than do either the instrumental record or numerical models.
- The paleorecord abounds with large and abrupt climate changes.
- The paleorecord includes numerous examples of the interaction between climate change in ecosystems and even societies.

Growing attention has been paid to the possible ramifications of potential future abrupt shifts beyond the range of variability on which planning and construction schemes are based and even outside the envelope of scenario future projections generated by climate models. The record of climate variability back through time reveals that such changes, often sudden, and sometimes persistent on decadal to century timescales. By the very nature of nonlinear systems, which are marked by sensitive dependence on initial conditions, such abrupt changes will remain largely unpredictable, no matter how sophisticated coupled, dynamical climate models may become. Thus, the best way to account for them is to develop scenarios based on past events.

Examples of abrupt climate events in the paleorecord are not limited to Greenland or glacial periods. Hydrological variability such as that shown in Figure 12.7, for example, includes sudden and high amplitude changes that
can be persistent for decades to centuries. Such variability lies well outside the range of instrumental records. Concepts of future environmental sustainability, water supply, and food security are limited and potentially dangerously short sighted if they fail to accommodate such evidence, especially given that these records are from regions that today experiences widespread catastrophic human casualties following on just a few consecutive years of drought. Climate variability and environmental change have played a major role in the welfare of human societies in the past, and all current indications are that their significance may be even greater in the future. The combination of high vulnerability to environmental hazards, such as floods and droughts, and continuing rapid population growth leave parts of the region under serious threat from increased variability, extreme events and future sea-level rise. Climate variability affects human welfare in a variety of ways and on a range of spatial and temporal scales. Equally, as land cover and agricultural practices change in response to evolving human needs there are important feedbacks to the climate system. The paleorecord holds vital clues as to how these interactions work. Improving the quality and long-term security of water supplies, enhancing agricultural productivity, and planning for the avoidance or mitigation of environmental hazards are all of high priority for human societies, and they all require a good understanding of past variability, human responses, and human–environment interactions.

References


Amman B, Oldfield F (2000) Biotic responses to rapid climatic changes around the Younger Dryas. Palaeogeogr Palaeoclimatol Palaeoecol 159: 175


13 Paleosols

Gregory Retallack

Abstract

Soils are known to be products of environmental factors such as climate, vegetation, topographic setting, parent material, and time for formation so that paleosols, or fossil soils, can potentially reveal changing environments of the past. Evidence from paleosols for past climate and vegetation in East Africa does not support traditional narratives of human evolution during a single transition from primeval forest to dry climate and open grassland. Instead, paleosols indicate climatic oscillations between wet and dry, and alternating expansion of woodland and grassland, since at least 18 Ma (million years ago). Acquisition of dry grassland adaptations such as thick enamel by 18 Ma, adducted hallux by 14 Ma, and cursorial legs by 1.8 Ma, alternated with woodland adaptations such as short stiff back by 16 Ma, erect stance by 6 Ma, and flat face by 3.5 Ma. Our ancestors survived profoundly changing climate and vegetation, with some adaptations lasting only to the next environmental shift, but others proving to be of lasting value.

13.1 Introduction

Our species, and its ancestors of millions of years ago, evolved on the soils of Africa (Darwin 1872; Fleagle 1998). Many of those soils have been eroded or altered beyond recognition by deep burial, but many paleosols are buried within floodplain, volcanic, lacustrine, alluvial plain, and cave deposits of Africa (Retallack 2001a). These paleosols provide much of the colorful banding and mottling seen in East African badlands and dongas, including many fossil hominoid localities (Figure 13.1). Many fossils of human ancestors come from paleosols (Retallack et al. 1995, 2002; Radosevich et al. 1997; Wynn 2004a, b), which are also records of past environments of our evolutionary antecedents (Figure 13.1). Modern soils are known to be products of environmental factors such as climate, vegetation, parent materials, topographic setting, and time for formation (Jenny 1941). These formative factors can be interpreted from fossiliferous paleosols to provide hitherto unavailable details of the habitats of fossil apes and humans.
The vegetation of fossil ape and human sites is central to long-standing theories for the evolution of upright stance. “The hands and arms could hardly have become perfect enough to have manufactured weapons, or to have hurled stones and spears with a true aim, ‘... so long as they were especially fitted to climbing trees’” (Darwin 1872). Other ideas are that grasslands selected for upright stance because of the need to be vigilant against predators (Dart 1926), to manipulate small seeds (Jolly 1970), to minimize exposure to the sun (Wheeler 1984), or to cover long distances with less energy by walking (Rodman and McHenry 1980) or running (Bramble and Lieberman 2004). Wooded grasslands and open woodlands are also plausible sites for evolution of upright stance from squat feeding on the ground (Kingdon 2003) or moving between scattered fruiting bushes (Sanford 2003). Alternatively, upright stance may have evolved in forests because it allowed erect-back climbing (Tuttle 1981), hands free to care for premature infants (Lovejoy 1981), phallic display to females (Tanner 1981), or intimidation displays to rivals (Jablonski and Chaplin 1993).
Paleosols are relevant to these questions because the fine root traces and crumb structure of grassland soils are distinct from the thick clayey subsurface horizons of both woodland and forest soils (Jenny 1941). Woodland and forest soils differ markedly in their clay minerals and chemical composition (Retallack 1997). Even the aquatic theory of human origins (Morgan 1982) can be evaluated from paleosols because mangal, littoral, lake margin, and streamside paleosols are distinguished by virtue of relict bedding and common burrows of crabs and clams (Retallack 2001).

Paleoclimate also is of interest as a selective pressure on hominoid evolution through drought and other hardships. Paleoclimate was also a primary control of past vegetation in which hominoids found food and shelter. Paleoclimatic shifts to drier climate and more open grassy vegetation have been held responsible for major evolutionary innovations in hominoids and bovids (Vrba 1999), as have changes in degree of climatic variability (Potts 1996). Soils of dry climate have calcareous nodules at a shallower depth than soils of humid climate (Retallack 2004) and also are less leached of cationic nutrients (Ca$^{2+}$, Mg$^{2+}$, K$^+$, Na$^+$) than humid-climate soils (Sheldon et al. 2002). Paleosols can thus provide paleoclimatic records at the very sites of early human ancestors rather than inferred from remote records of deep-sea cores (de Menocal 2004).

This review emphasizes climate and vegetation, but other soil forming factors of parent material, topographic position, and duration of soil formation can also be inferred from paleosols. Highly calcareous and saline carbonatite volcanics are an unusual parent material of many African hominid sites (Retallack 1991a), fortunate because of their remarkable preservative effects for fossil bones, seeds, and insects (Retallack et al. 1995). Many East African paleosols preserve a record of well-drained fluvial terraces, infrequently flooded, and some volcanic apron paleosols represent foothills environments (Retallack 1991a), thus revealing environments beyond the usual lowland constraints of sedimentary environments. Degree of soil development can also be used to infer duration of paleosol formation and rates of sediment accumulation, with implications for the geochronology of ape and human ancestor sites (Retallack et al. 1995, 2002).

### 13.2 Recognition of paleosols

Paleosols are often distinctive and striking bands of red clay (Bt horizon), calcareous nodules (Bk horizon), or coal (O horizon) in sedimentary and volcanic sequences (Figure 13.2). Three general classes of observations are especially helpful in paleosol recognition: root traces, soil horizons, and soil texture (Retallack 1997).
Kenyan Miocene paleosols have been given field names using local Luo and Turkana languages. These pedotypes are objective field mapping units for paleosols: their interpretation and classification requires laboratory study.

![Figure 13.2](image-url)
Root traces are the most diagnostic evidence of paleosols and sometimes are obvious because cemented and erosion-resistant (Kabisa pedotype of Figure 13.2). Difficulties arise in recognition of root traces because they are often replaced by other minerals and ramify in three dimensions in such a way that one rock face reveals little of the overall pattern. Few fossil roots are carbonaceous or reveal histological structures like fossil wood (Retallack 1997). The original root has commonly rotted out, and the hole it occupied is filled with claystone or siltstone, or encrusted with iron oxide or calcium carbonate. Drab haloed root traces are very distinct, green gray mottles, in reddish paleosol matrix, formed during early burial chemical reduction by microbes fueled by consumption of root organic matter (Retallack 1991b). In all these cases, root traces are truncated at the surface of the paleosol, and branch and taper downward. These features distinguish root traces from most kinds of burrows in soils, although the relationship between burrows and roots can be complex. Roots may preferentially follow soft fill of burrows rather than hard soil matrix, and burrows may congregate around roots on which the burrowing animals fed (Retallack 1991a).

Soil horizons develop through thousands of years whereas sedimentary beds are deposited in days. Unlike sedimentary beds, which have sharp bottoms and usually sharp tops as well, paleosols have a sharp top, representing the ancient land surface, but gradational lower contacts (Retallack 2001a). Sedimentary beds also include a variety of sedimentary structures, such as lamination, cross bedding, and ripple marks (as in Tek pedotype of Figure 13.2), whereas soil horizons develop with obliteration of these original features (Tut pedotype of Figure 13.2). Similarly, soil formation progressively destroys the original crystalline structure of volcanic or granitic parent materials (Retallack 1991a). In dry climate soils (Aridisols), primary sedimentary or volcanic structures are obscured at first by filaments and soft, small carbonate masses, then large, hard, carbonate nodules (calcic or Bk horizon of Chogo pedotype in Figure 13.2), and finally thick carbonate layers (petrocalcic or K horizon of Soil Survey Staff 2000). In sod-grassland soils (Mollisols), primary lamination and crystalline structure is broken up by fine roots and replaced by dark, fecal pellets of earthworms to create a crumb-textured, organic surface horizon (mollic epipedon of Dite, Chogo, Yom, and Onuria pedotypes of Figure 13.2). A variety of other kinds of soil horizons are recognized and important to soil classification (Retallack 1997, 2001a; Soil Survey Staff 2000).

Soil structure also develops within soil horizons and is very distinct from sedimentary bedding and igneous crystalline texture. The fundamental elements of soil structure are modified cracks and other surfaces (cutans), and the clods they define (peds). Cutans include clay skins (argillans) lining cracks in the soil.
and rusty weathering rinds (sesquans) around clods and pebbles in soil (Retallack 2001a). Peds have a variety of shapes: lenticular in swelling-clay soils (Vertisols: Aberegaiya pedotype of Figure 13.2), blocky subangular in fertile forest soils (Alfisols: Tut pedotype of Figure 13.2), and crumb shaped (small and ellipsoidal) in grassland soils (Mollisols: Dite, Chogo, Yom, and Onuria pedotypes of Figure 13.2). Although cracks and other voids are not preserved in paleosols due to compaction by overburden, peds, and cutans are common and conspicuous (Retallack 1991a). Other soil structures less diagnostic of soils include concretions, nodules, and crystals (Retallack 2001a).

### 13.3 Methods for the study of paleosols

Just as soil individuals (pedons) are studied as soil columns in soil pits, paleosols are studied in columnar stratigraphic sections of the sort also used in sedimentology and stratigraphy (Figure 13.3). Grain size is emphasized because it is important to soil formation, as weathering transforms sand and silt grains to clay. A graphical representation of grain size profiles conveys important information on the abruptness of horizon transitions. Color from a Munsell chart should also be represented, as redness denotes the degree of chemical oxidation and drainage of soils and paleosols. Calcareousness determined by relative effervescence with dilute hydrochloric acid also is important as a guide to chemical leaching and soil nutrient status (Retallack 1997).

Laboratory studies of paleosols do not employ all the same techniques used in soil science because some important soil measures, such as base saturation, are altered upon burial of soils (Retallack 1991b). Petrographic thin sections are especially useful for revealing soil microfabrics, and the point counting of thin sections furnishes estimates of changes in grain size and mineral composition of paleosols. For example, increased subsurface clayeyness can be used to recognize diagnostic horizons (argillic horizon) for forest soils (Alfisols and Ultisols), whereas traces of nutrient-rich minerals, such as calcite and feldspar, distinguish fertile forest soils (Alfisols) from infertile forest soils (Ultisols: Retallack 1997).

Chemical analyses also are useful in characterizing and classifying paleosols, especially molar ratios designed to gauge the progress of common soil-forming chemical reactions. The hydrolysis reaction common in silicate weathering leaches cationic bases ($\text{Ca}^{2+}$, $\text{Mg}^{2+}$, $\text{K}^+$, $\text{Na}^+$) from host minerals, such as feldspar, to create clay (Al rich) and is thus indicated by high ratios of alumina/bases. An alumina/base ratio higher than 2 is a good proxy for the transition from fertile forest soils (Alfisols) to infertile forest soils (Ultisols). Soda/potash molar ratios in excess of 1 indicate unusually salty soils. Ferrous/ferric molar ratios in
Figure 13.3
Petrographic and chemical data for the Tek paleosol from the 18 Ma Hiwegi Formation of Rusinga Island, Kenya [data from Retallack et al. (1995)]
excess of 1 indicate well-drained soils (Retallack 1997). By these criteria, the petrographic and chemical data on the 18 Ma Tek paleosol from Rusinga Island Kenya (Figure 13.3) indicate a fertile Inceptisol that was nonsaline and well-drained. These data allow identification of analogous modern soils (Retallack et al. 1995) and refine understanding of the ancient landscape and its ecosystem (Figure 13.4).

Figure 13.4
Reconstruction of Tek paleosol from the 18 Ma Hiwega Formation of Rusinga Island, Kenya [data from Retallack et al. (1995)]
Carbon isotopic compositions of paleosol carbonate were at first thought to be useful indicators of grasslands because most tropical grasses have a C$_4$ photosynthetic pathway which creates isotopically heavy carbon (Cerling 1992). The most prominent failure of this technique was application to the Middle Miocene (14.4 Ma) locality of Fort Ternan, with paleosol carbonate isotopically like rain forest (Cerling et al. 1997a), but fossil soils, grasses, trees, and antelope like those of a mosaic of wooded grassland and grassy woodland (Retallack 1991a, 1992; Koch 1998; Turner and Antón 2004). Subsequently it was found that even tropical grasses used the C$_3$ photosynthetic pathway until about 7 Ma or younger (Cerling et al. 1997b; Fox and Koch 2003). The quality of graze (C$_3$ more nutritious than C$_4$) can be assessed by isotopic studies of teeth and paleosols, but the question of grass or shrub diet is better assessed from mammalian tooth microwear, hypsodonty, and cursoriality (MacFadden 2000). The advent of C$_4$ grasses within tropical grasslands is most likely related to declining Late Miocene atmospheric CO$_2$ content (Cerling et al. 1997a). Another failure of carbon isotopes to indicate past vegetation is Sike’s (1994) forest interpretation of the paleosol at Olduvai fossil locality FLK yielding *Australopithecus boisei*. This paleosol with relict bedding, zeolites, little clay, and shallow carbonate is unlike forest soils, and probably supported salt-tolerant, lake-margin shrubs (Retallack 2001a), which have a similar C$_3$ isotopic value to forest (Sikes 1994). Isotopic values of carbon and oxygen in paleosols and animals are controlled by so many factors that biotic and pedogenic constraints are needed (Koch 1998). Carbon isotopic studies of paleosols are now more useful for assessing atmospheric CO$_2$ from carbonate and organic isotopic offsets (Ekart et al. 1999) and soil productivity from isotopic depth functions (Yapp and Poths 1994). Carbon isotopic depth profiles of paleosols also provide new insights into carbon cycling within different kinds of ancient ecosystems. Grassland paleosols (Chogo and Onuria pedotype of Figure 13.5) show more effective humification at the surface (higher $\delta^{13}$C values) than woodland soils (Tut of Figure 13.5), and swelling-clay paleosols have flat carbon isotopic profiles due to soil mixing (Chido of Figure 13.5). Preservation of such carbon-cycling signatures known from modern soils within different pedotypes gives additional evidence for paleosol classification and interpretation (Bestland and Krull 1999).

### 13.4 Paleosols as proxies of paleoprecipitation

Climatic zonation of soils was a key element in the Russian origins of soil science (Jenny 1941), and a variety of relationships between particular soil features and climatic variables can be applied to East African paleosols in order to reconstruct...
paleoclimate. For example, depth to carbonate horizon \((D\text{ in cm})\) is related to mean annual precipitation \((P\text{ in mm})\) by formula (1) below (from Retallack 2005). This depth can be corrected for compaction due to overlying sediment using geological estimates of overburden and standard formulae (Sheldon and Retallack 2001). Also related to mean annual precipitation \((P)\) is nutrient base content \((C = \frac{Al_2O_3}{Al_2O_3 + CaO + MgO + Na_2O})\) in mol) of soil Bt horizons by formula (2) below (Sheldon et al. 2002).

\[
P = 137.24 + 6.45D - 0.013D^2 \quad R^2 = 0.52, \text{ S.E. } = \pm 147 \text{ mm} \quad (1)
\]

\[
P = 221.12 \cdot e^{0.197C} \quad R^2 = 0.72, \text{ S.E. } = \pm 182 \text{ mm} \quad (2)
\]

Chemical weathering also alters the mineral content of soils, especially their clay minerals, which begin as smectites and then lose cationic bases with further chemical weathering to become kaolinite (Retallack 2001). This indication of paleoprecipitation works best with noncalcareous soils, which are found in climates receiving more than 1,000-mm mean annual precipitation (Retallack 2004). In East Africa today, smectite is dominant in soils receiving less than 1,200-mm mean annual precipitation, and kaolinite dominant in wetter climates.
Thus noncalcareous, smectitic soils define a limited paleoclimatic window of 1,000–1,200-mm mean annual precipitation.

My new compilation of Kenyan paleoprecipitation over the past 20 million years (Figure 13.6b) includes previously published data on African depth to Bk (Wynn 2001, 2004a, b; Wynn and Retallack 2002; Retallack 2001b; Retallack et al. 2002), and paleosol chemical (Retallack et al. 1995, 2002; Thackray 1989; Bestland 1990; Retallack 1991a; Wynn and Retallack 2002) and clay mineral composition (Retallack 1991a, Behrensmeyer et al. 2002), as well as published inferences from size and shape of fossil leaves (Jacobs 2002). This compilation is limited to data from around Lake Victoria for the early-middle Miocene, the Tugen Hills for the mid-late Miocene and the Turkana Basin for the Miocene to Quaternary. The geological time scale is from radiometric dating of these various fossil primate sites (Deino et al. 1990; Retallack 1991a, Jacobs and Deino 1996; Behrensmeyer et al. 2002; Hill et al. 2002).

These new data reveal not just one Neogene aridification event at about 7 Ma, as has long been implied by the “Tertiary pluvial hypothesis” (Leakey 1952), the “Miocene lake hypothesis” (Kent 1944), the “Miocene rain forest hypothesis” (Andrews and Van Couvering 1975; Andrews 1996), and the “Late Miocene grassland hypothesis” (Cerling 1992; Cerling et al. 1997a, b). These theories had already been discredited by discovery of Miocene desert dunes, shrubland snails, alkaline lakes, open-country grasses, grazing mammals, and grassland paleosols in East Africa (Pickford 1986a, 2002a; Retallack et al. 1990, 2002).

Instead the data (Figure 13.6) reveal a Neogene paleoclimatic roller coaster of at least nine dry spells with intervening wet periods, of which humidity spikes at 16 and 13 Ma were the wettest of the last 20 million years. This new paleoprecipitation curve is similar to paleotemperature variations for Africa inferred from north–south oscillation through time of Ethiopian and Palearctic biogeographic realms (Pickford 2002a). These new data are also similar to foraminiferal oxygen isotope curves from the deep sea (Zachos et al. 2001), commonly used as a basis for evaluating human evolution in Africa (de Menocal 2004), but the match is not precise (Figure 13.6b and c). A general trend of extreme and volatile middle Miocene values, but subdued late Miocene to Quaternary values, is evident from both isotopic and paleosol data. The paleosol record reveals much greater variation in rainfall than would be inferred from carbon isotopic values of marine foraminifera, which are damped by global oceanic mixing with time lags of several thousand years. More profound damping is seen in oxygen isotopic values of marine foraminifera, which show a long-term increase unlike local rainfall and foraminiferal carbon records. This increase is plotted on reversed axes in Figure 13.6d because it has been interpreted as a long-term temperature
Figure 13.6
A 20 million year record of vegetation (a) and paleoprecipitation (b) from Kenya, compared with carbon (c) and oxygen (d) isotopic composition of marine foraminifera. Paleoprecipitation data from paleosols (b) is from depth to carbonate (open ellipses), clay mineral (diamonds), and chemical composition (squares) after Retallack (1991a), Retallack et al. (1995, 2002), Wynn (2001, 2004a, b), Wynn and Retallack (2001). Paleobotanical estimates from Jacobs (2002) and Jacobs and Deino (1996). Modern vegetation precipitation limits are from Anhuf et al. (1999).
decline (Zachos et al. 2001), but part of this long-term trend is due not just to temperature but to water recycling with plate tectonics (Veizer et al. 2001). The global oxygen isotope record also shows an increase after 3 Ma due to continental icecap sequestration of isotopically light oxygen, in addition to temperature effects (Zachos et al. 2001). Despite these problems, the East African paleosol record and global isotopic records present a very different concept of climatic variation experienced by our distant ancestors than the past idea of a seminal Late Miocene climatic event. Instead of a single origin of humanity at a turning point of environmental change, the new record implies rather that our lineage responded to a gauntlet of changing conditions with a variety of adaptations (Table 13.1), as discussed later.

Table 13.1
Geological age of African climatic events, selected adaptations, and hominoid diversity (D), origination (O), and extinction (E)

<table>
<thead>
<tr>
<th>Age (Ma)</th>
<th>Hominoid adaptations and extinctions</th>
<th>D</th>
<th>O</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>20.2 dry</td>
<td>Robust mandible for hard food (<em>Rangwapithecus</em>)</td>
<td>10</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>19.1 wet</td>
<td>Low cusp molars for folivory (<em>Nyanzapithecus</em>)</td>
<td>5</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>17.7 dry</td>
<td>Thick enamel for hard food (<em>Afropithecus</em>)</td>
<td>6</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>16.1 very wet</td>
<td>Short back for suspension (<em>Morotopithecus</em>)</td>
<td>4</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>14.9 very dry</td>
<td>Adducted hallux for ground walking (<em>Kenyapithecus</em>)</td>
<td>7</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>12.6 very wet</td>
<td>Thin enamel molars for soft food (<em>Otavipithecus</em>)</td>
<td>2</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>10.7 dry</td>
<td>Large size for ground feeding (<em>Samburupithecus</em>)</td>
<td>3</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>8.6 wet</td>
<td>Ape extinction with monkey radiation (<em>Microcolobus</em>)</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>7.5 very dry</td>
<td>Knuckle walking for ground (<em>Pan-Gorilla ancestors</em>)</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>6.8 wet</td>
<td>Upright stance for nest provisioning (<em>Orrorin</em>)</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5.4 dry</td>
<td>Small incisiform canines for hard food (<em>Ardipithecus</em>)</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>4.2 wet</td>
<td>Flat face for stereoscopic vision (<em>Kenyanthropus</em>)</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>2.5 dry</td>
<td>Large molars for hard food (<em>Paranthropus</em>)</td>
<td>6</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>2.1 wet</td>
<td>Small molars for soft food (<em>Homo habilis</em>)</td>
<td>3</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>1.8 dry</td>
<td>Long legs for endurance running (<em>Homo ergaster</em>)</td>
<td>5</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>1.7 wet</td>
<td>Occipital bun for competition (<em>Homo erectus</em>)</td>
<td>4</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>1.0 dry</td>
<td>Globular brain for generalist roles (<em>Homo antecessor</em>)</td>
<td>4</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>0.1 wet</td>
<td>Magdalenian tools and culture (<em>Homo sapiens</em>)</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
</tbody>
</table>

13.5 Paleosols as trace fossils of ecosystems

*Australopithecus afarensis* is known from body fossils, such as the partial skeleton “Lucy” (Johanson et al. 1982), as well as from trace fossils, such as the footprints of Laetoli (Leakey and Harris 1987). The soils of *A. afarensis* also are known, especially at the “first family” site near Hadar, Ethiopia (Radosevich et al. 1992). Here a troop of at least 13 individuals, young and old, died, rotted, and were
partially disarticulated, before being interred in flood deposits on a crumb-structured soil of grassy streamside woodland (Figure 13.7). The paleosol is not only a matrix to the bones but a trace fossil of their ecosystem. Furthermore, paleosols by definition are in the very place they formed, not redeposited. Unlike the skeleton of “Lucy” found in the sandstone of a former river channel (Johanson et al. 1982), and thus transported some distance from its natural habitat, the first family was found where it died and had lived (Radosevich et al. 1992). Thus, paleosols give a finer resolution of primate paleoenvironments in time (Figure 13.6) as well as space (Figure 13.7).

The various paleosols containing Miocene ape fossils in southwest Kenya can also be used to constrain their habitats (Figure 13.8). The fragmentary and weathered nature of most of these fossils is evidence that they accumulated through natural processes of death and decay on the paleosols in which they are found (Pickford 1986a). The great diversity of fossil apes in this region (Gommery et al. 2002; Harrison 2002; Ward and Duren 2002) is in contrast to the low diversity of great apes today (Fleagle 1998), leading to the idea that Miocene apes, defined from apelike dentition, were ecologically more like both
Figure 13.8
Paleosols of Miocene apes from southwestern Kenya [data from Retallack (1991a) with taxonomy after Harrison (2002), Retallack et al. (2002), Ward and Duren (2002)]
apes and monkeys today (Andrews 1996). Analysis of their occurrence in paleosols shows that there was some ecological separation of different species to different soil types, but still high diversity within a soil type (Figure 13.8). In the 20 Ma sites of Koru and Songhor, for example, one taxon (Ugandapithecus) shows little habitat specificity through a variety of tropical dry forest habitats, but small taxa (Kalepithecus, Mabokopithecus) are in upland soils and larger taxa (Proconsul, Rangwapithecus, Dendropithecus, Nyanzapithecus) remained in lowland forests closer to water. One paleosol type (Kiewo pedotype) has as many as six taxa: three likely suspensory feeders (Limnopithecus, Dendropithecus, Ugandapithecus from small to large) and three likely overbranch feeders (Nyanzapithecus, Proconsul, Rangwapithecus, from small to large). The contrasting sizes and other differences between these taxa suggest niche partitioning of forest canopy tiers.

Diverse sympatric catarrhine communities persisted into the dry woodland landscapes of Rusinga Island at 17.8 Ma, when paleosols with the crumb peds and iron–manganese nodules of dambo grasslands (Yom pedotype) appear, but are rare and barren of primate fossils (Retallack et al. 1995). Other evidence for grasslands of about the same age are abundant bunch grasses at the Ugandan fossil site of Bukwa (Pickford 2002b). Yom paleosols of dambo grassland are much more common by 14.7 Ma on Maboko Island (Retallack et al. 2002), where they contain abundant vervet-like monkeys (Victoriapithecus: note change of scale for this exceptional collection in Figure 13.8). These seasonally inundated grasslands of dry climates were not encouraging to fossil apes, which were more common in riparian woodlands (Nyanzapithecus, Limnopithecus, Mabokopithecus and Simiolus of Dhero paleosols). More wide ranging was Kenyapithecus, found in both riparian woodland (Dhero) and nyika shrubland (Ratong), which it exploited more effectively than other apes because of its thick enameled, large molars useful for tough foods (Martin 1985) and its macaque-like limbs and feet ( McCrossin et al. 1998). A similar pattern of wide ranging Kenyapithecus and forest-dependent other apes (Oreopithecus, Simiolus, Proconsul) persisted in grassland mosaics of Fort Ternan and Kapsibor at 14.4 Ma, when well-drained short-grass, wooded grassland was widespread. The appearance of grasslands so encouraging for victoriapithecine ancestors of vervets and colobines, was not so encouraging to apes, which remained rare components of the fossil fauna.

Reconstruction of rainfall from paleosols implies also vegetation belts (Figure 13.6a), by comparison with Holocene climatic ranges of plant formations (Anhuf et al. 1999). There was rainforest in central Africa during the past 20 million years as indicated by rare finds of fossil plants (Bancroft 1932, 1933), but evidence of rain forest has not yet been found in the East African areas of hominoid fossils (Retallack 1991a, Jacobs 2002). These interpretations (Figure 13.6a) are well in accord with indications of vegetation from paleosol classification, profile
form and root traces (Retallack 1991a), as evidence that the climatic range of
most vegetation types did not change over the past 20 million years.

An exception is the evolution of grasslands, which expanded their climatic
range to displace extinct kinds of woodlands (Figures 13.6 and 13.9). There is
not yet any East African evidence of grasslands before 17.8 Ma, when crumb-
textured, brown, simple (A-Bk) profiles of dambo were rare at Rusinga Island
(Retallack et al. 1995) and bunchgrasses grew luxuriantly at Bukwa (Pickford
2002b). Well-drained, short-grass, sod-grasslands were widespread by 14.4 Ma
(Retallack 1991a, Retallack et al. 2002) and well-drained, tall-grass, sod-grasslands
expanded their climatic range considerably by 7 Ma (Wynn 2004a, b). Grasslands
were a newly coevolved ecosystem of the Cenozoic, with grasses uniquely suited
to grazing by virtue of their intercalary meristems, modular growth, basal tiller-
ing, and sod formation, and grazers uniquely suited to coarse grassy fodder by
virtue of their wide muzzles, hypsodont teeth, and hard hooves (Retallack 2001a).
A world without grasslands was transformed over some 20 million years to a Plio-
Pleistocene world with grassland covering at least a quarter of the land surface.
Holocene humans spread grassy agroecosystems to almost all parts of the world
(Retallack 2001a). Neogene expansion of grasslands within the paleoclimatic belt
roughly defined by the 300–750 mm per annum isohyet enabled grasslands to
capture the planetary modal rainfall belt and most fertile soils, with consequences
for global change including a significant contribution to global cooling (Retallack
2001b).

Before the expansion of the grasslands, an extinct woody vegetation occupied
their climatic range (Figure 13.6). I call these extinct dry woodlands pori
(Table 13.2), from a Hadza word for bush (Woodburn 1968). A good example
of a pori ecosystem is the Tek paleosol of Rusinga Island (Figures 13.3
and 13.4), which has yielded fossil primates and other mammals, snails, and plants
(Pickford 1995; Retallack et al. 1995). Other examples of pori ecosystems include
Tut, Choka, and Kwar pedotypes of Songhor and Koru dated at 20 Ma (Retallack
1991a). From the soil perspective, these paleosols have no clear modern analog
because they are red and clayey, with large root traces and blocky structure like
woodland soils, yet have shallow calcareous horizons like those found in modern
African semiarid to subhumid grassland soils. Modern African soils with such
shallow carbonate have very different crumb structure, fine root traces, and dark
brown organic-rich surface horizons from abundant grasses.

From the paleoanthropological perspective, these ancient communities have
no modern analogs because they have so many fossil hominoids, as many as six
species in the Kwar pedotype (Figure 13.8). No community has so many species
of hominoids today. Nor do modern hominoids live in such dry climates. Mt. Assirik in Senegal with 956 (854–1224) mm mean annual precipitation is
the driest climate with chimpanzees (Kappelman 1993), although Kingdon (2003) gives anecdotes of chimpanzees in wooded grassland. It is now clear that Miocene apes filled a variety of niches like those today filled by vervets, baboons, and colobines as well as apes (Retallack et al. 2002). Pori ecosystems such as the Kwar paleosol at Koru (20 Ma) also show peculiar associations of other mammals, including a mix of dry climate taxa, such as mole rats (*Bathyergoides*), with wet climate taxa such as flying squirrels (*Paranomalous*), giant elephant shrews (*Miorynchocyon clarki*), tenrecs (*Protenrec tricuspis*), golden moles (*Prochrysochloris miocaenicus*), and chevrotains (*Doratherium songhorense*; Retallack 1991a). Similarly the Tut and Choka paleosols at Songhor (20 Ma) and Tek paleosols on Rusinga Island (17.8 Ma) have wet climate flying squirrels and tenrecs as well as dry climate mole rats and spring hares (Retallack 1991a; Retallack et al. 1995). These nonanalog combinations of fossil mammals can be explained by a theory of evolutionary replacement of pori with grassland within semiarid to subhumid regions. Before the advent of grasslands, woody vegetation became smaller in stature and biomass from wet to dry regions. This continuum was disrupted as grasslands evolved to usurp the climatic range of pori woodland. Grasslands expanded their range to create a biogeographic divide between Nyika shrubland and miombo woodland (Figure 13.9).

Fossil primates of East Africa not only coped with changing mixes of animals but with changing climate and vegetation (Figure 13.6). Wynn (2004b) has introduced the concept of evolutionary entropy to explain effects of climate and vegetation change on hominoid diversity. Climatically dry episodes encouraged grassland mosaic environments with a more varied landscape of open grassland and local woodland, and thus greater landscape disorder or negentropy. Wet episodes of forest vegetation presented more uniform landscapes of higher entropy. My compilation of hominoid diversity, originations, and extinctions (Table 13.1, Figure 13.10) supports the view that dry episodes correspond with diverse primates, whereas wet episodes lead to extinctions, particularly of specialized arid-adapted taxa. The concept of ecosystem entropy in hominoid evolution is similar in some respects to Vrba’s (1999) “turnover pulse hypothesis,” but ecosystem entropy presents diffuse and long-term selection pressures, rather than episodic crises or “turnover pulses.” Recent compilations of mammalian
CLIMATIC RANGE OF LIVING EAST AFRICAN MAMMALS

- apes (Pan troglodytes, Gorilla gorilla)
- monkeys (Colobus badius, Colobus polycomos, Galago demidovii, Galago senegalensis)
- bush babies (Papio cynocephalus)
- vervet baboon (Cercopithecus aethiops)
- tenrecs (Potamogale velox, Microtamias ruwenzori)
- golden mole (Chrysoschillosternum stuhlmanni, Rhynchochus cimei)
- mole shrews (Elephantulus rufescens)
- mole rats (Cryptomys ochraceocineus, Cryptomys hottentotus, Heliophobius argenteocineus, Heterocephalus glaber)
- flying squirrels (Idius renkeri, Anomalurus derbianus)
- spring hare (Pedetes capensis)
- cane rats (Thryonomys swinderianus, Thryonomys gregorianus)
- chevrotain (Hyemoschus aquaticus)

Plio-Pleistocene
(0.5-7 Ma)
- advent of tall sod grassland

Middle-Late Miocene
(7-18 Ma)
- advent of short sod grassland

Oligocene-Early Miocene
(18-33 Ma)
- desert rangeland and pori woodland

SELECTED NON-ANALOG FOSSILIFEROUS PALEOSOLS

TEK paleosol Rusinga Island 18 Ma
Choka paleosol Songhor 20 Ma
Tut paleosol Songhor 20 Ma
data from the Turkana region do not show such crises (Bobe et al. 2002), revealing instead an oscillating diversity compatible with less synchronized selection by ecosystem entropy.

A major caveat for such theories is the generally inferior fossil record of climatic wet phases because their soils and sediments are noncalcareous and so not favorable to the preservation of bone (Retallack 1998). We still have no primate fossil record from paleoclimatic wet phases of the early Miocene, but there are discoveries of wet climate human ancestors from 13 Ma (Hill et al. 2002), 6 Ma (Brunet et al. 2002; Galik et al. 2004) and 4–3 Ma (Carroll 2003). The soil-taphonomic bias against wet climate fossils makes the search difficult, not impossible (Peterhans 1993).

Each fluctuation in climate and vegetation presented new crises and opportunities to primates. My own correlation of climatic events with critical adaptations (Table 13.1) is only an outline of a new research agenda, to be fleshed out with further studies of the critical intervals. The late Miocene paleosols and primate fossils of the Tugen Hills, for example, remain very poorly known compared with those of the Lake Victoria and Turkana basins. Nevertheless, there are general themes apparent from this compilation. We did not evolve from apes in one seminal event, but by a protracted process of growth and pruning of our evolutionary tree. Some specialized features such as procumbent incisors at 18 Ma evolved in dry grassy woodlands, but did not survive succeeding forest expansions (McCrossin and Benefit 1997). Some specialized features such as long arms

<table>
<thead>
<tr>
<th>Feature</th>
<th>Pori Dry woodland</th>
<th>Miombo Dry woodland</th>
<th>Nyika Dry bushland</th>
<th>Savanna Wooded grassland</th>
</tr>
</thead>
<tbody>
<tr>
<td>Key genera</td>
<td>Celtis</td>
<td>Brachystegia</td>
<td>Acacia</td>
<td>Combretum</td>
</tr>
<tr>
<td>Floral origins</td>
<td>Zambezi</td>
<td>Zambezi</td>
<td>Zambezi</td>
<td>Eurasian</td>
</tr>
<tr>
<td>Spinosity</td>
<td>Unarmed</td>
<td>Unarmed</td>
<td>Spinose</td>
<td>Spinose</td>
</tr>
<tr>
<td>Leaf set</td>
<td>Semideciduous</td>
<td>Deciduous</td>
<td>Deciduous</td>
<td>Deciduous</td>
</tr>
<tr>
<td>Fruit size</td>
<td>Large</td>
<td>Large</td>
<td>Small</td>
<td>Small</td>
</tr>
<tr>
<td>Snails</td>
<td>Cerasia</td>
<td>Limicolaria</td>
<td>Achatina</td>
<td>Pupoides</td>
</tr>
<tr>
<td>Snail origins</td>
<td>Somalian</td>
<td>Somalian</td>
<td>Somalian</td>
<td>Somalian</td>
</tr>
<tr>
<td>Mammals</td>
<td>Apes, rodents</td>
<td>Antelope</td>
<td>Antelope</td>
<td>Antelope</td>
</tr>
<tr>
<td>Ungulates</td>
<td>Walangania</td>
<td>Aepyceros</td>
<td>Tragelaphus</td>
<td>Connochaetes</td>
</tr>
<tr>
<td>Primates</td>
<td>Proconsul</td>
<td>Cercopithecus</td>
<td>Papio</td>
<td>Papio</td>
</tr>
<tr>
<td>Mammal origin</td>
<td>Zambezi</td>
<td>Zambezi</td>
<td>Zambezi</td>
<td>Eurasian</td>
</tr>
<tr>
<td>Fire frequency</td>
<td>Low</td>
<td>High</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Soil organics</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Soil fertility</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Soil type</td>
<td>Alfisol</td>
<td>Oxisol, Vertisol</td>
<td>Aridisol</td>
<td>Mollisol, Vertisol</td>
</tr>
<tr>
<td>Parent material</td>
<td>Volcanic</td>
<td>Granitic</td>
<td>Granitic</td>
<td>Volcanic</td>
</tr>
</tbody>
</table>

Table 13.2
Comparison of extinct pori woodland with extant East African vegetation

<table>
<thead>
<tr>
<th>Feature</th>
<th>Pori Dry woodland</th>
<th>Miombo Dry woodland</th>
<th>Nyika Dry bushland</th>
<th>Savanna Wooded grassland</th>
</tr>
</thead>
<tbody>
<tr>
<td>Key genera</td>
<td>Celtis</td>
<td>Brachystegia</td>
<td>Acacia</td>
<td>Combretum</td>
</tr>
<tr>
<td>Floral origins</td>
<td>Zambezi</td>
<td>Zambezi</td>
<td>Zambezi</td>
<td>Eurasian</td>
</tr>
<tr>
<td>Spinosity</td>
<td>Unarmed</td>
<td>Unarmed</td>
<td>Spinose</td>
<td>Spinose</td>
</tr>
<tr>
<td>Leaf set</td>
<td>Semideciduous</td>
<td>Deciduous</td>
<td>Deciduous</td>
<td>Deciduous</td>
</tr>
<tr>
<td>Fruit size</td>
<td>Large</td>
<td>Large</td>
<td>Small</td>
<td>Small</td>
</tr>
<tr>
<td>Snails</td>
<td>Cerasia</td>
<td>Limicolaria</td>
<td>Achatina</td>
<td>Pupoides</td>
</tr>
<tr>
<td>Snail origins</td>
<td>Somalian</td>
<td>Somalian</td>
<td>Somalian</td>
<td>Somalian</td>
</tr>
<tr>
<td>Mammals</td>
<td>Apes, rodents</td>
<td>Antelope</td>
<td>Antelope</td>
<td>Antelope</td>
</tr>
<tr>
<td>Ungulates</td>
<td>Walangania</td>
<td>Aepyceros</td>
<td>Tragelaphus</td>
<td>Connochaetes</td>
</tr>
<tr>
<td>Primates</td>
<td>Proconsul</td>
<td>Cercopithecus</td>
<td>Papio</td>
<td>Papio</td>
</tr>
<tr>
<td>Mammal origin</td>
<td>Zambezi</td>
<td>Zambezi</td>
<td>Zambezi</td>
<td>Eurasian</td>
</tr>
<tr>
<td>Fire frequency</td>
<td>Low</td>
<td>High</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Soil organics</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Soil fertility</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Soil type</td>
<td>Alfisol</td>
<td>Oxisol, Vertisol</td>
<td>Aridisol</td>
<td>Mollisol, Vertisol</td>
</tr>
<tr>
<td>Parent material</td>
<td>Volcanic</td>
<td>Granitic</td>
<td>Granitic</td>
<td>Volcanic</td>
</tr>
</tbody>
</table>
by 20 Ma for suspensory locomotion in forests did not persist through succeeding grassland expansions (Harrison 2002). Other forest adaptations such as a short-stiff back by 16 Ma (Pickford et al. 1999), erect stance by 6 Ma (Senut et al. 2001; Galik et al. 2004), and flat face by 3.5 Ma (Leakey et al. 2001) proved advantageous in the long term, just as did grassland adaptations, such as thick enamel by 18 Ma (Martin 1985; McCrossin and Benefit 1997), adducted hallux by 14.7 Ma (McCrossin et al. 1998), and long legs for endurance running by 1.8 Ma (Bramble and Lieberman 2004). Although each of these ideas could be debated individually, the general concept of human evolution as a generalist path through a gauntlet of environmental challenges (Potts 1996) is increasingly supported by a burgeoning fossil record (Carroll 2003). There will always be a need for dating and finding more human ancestor fossils, but paleosols now provide new evidence of evolutionary selection pressures with high temporal and spatial resolution.

Past hypotheses of a Miocene pluvial, lake and rain forest (Kent 1944; Leakey 1952; Andrews and Van Couvering 1976; Andrews 1996) and late Miocene grassland (Cerling 1992; Cerling et al. 1997a, b) find, as already reviewed, a counterpart in long-standing theories linking late Miocene evolution of human upright stance or large brains with hunting prowess (Darwin 1872), vigilance against predators (Dart 1926), manipulation of small seeds (Jolly 1970), minimization of sun exposure (Wheeler 1984), long-distance walking (Rodman and
McHenry 1980) or running (Bramble and Lieberman 2004), squat feeding on the ground (Kingdon 2003), or moving between scattered fruiting bushes (Sanford 2003). Forest explanations of upright stance allowing erect-back climbing (Tuttle 1974), hands free to care for premature infants (Lovejoy 1981), phallic display to females (Tanner 1981), or intimidation displays to rivals (Jablonski and Chaplin 1993) move the event back into the “Miocene rain forest” (of Andrews and Van Couvering 1975; Andrews 1996), for which there is little evidence at hominoid sites in East Africa (Figure 13.6a). All these views can be reassessed in light of the improved record of East African paleosols, which suggests that there were many alternating habitats in East Africa, not just one seminal environmental shift. Darwin’s (1872) idea that erect stance was linked to tool use and brain expansion has been out of favor since the discovery of “Lucy,” when it became clear that erect stance preceded tool use and brain expansion by millions of years (Johanson et al. 1982). Erect stance now appears to have occurred in wooded habitats by 6 Ma (Pickford and Senut 2001; Vignaud et al. 2002), perhaps selected by the use of hands in nest provisioning (Lovejoy 1981). We are a mosaic of a complex evolutionary history and no longer need settle for simple or single allegories of human evolution.

13.6 Conclusions

There is a copious and informative fossil record of soils at most of the fossil ape and human ancestor sites in Africa, and study of these paleosols is now giving important insights into the long evolutionary career of our ancestors. The primate evolutionary radiation of the Neogene has been a long saga of changing habitats and adaptations. The fossil record of soils now allows us to address its complexity on a scale appropriate to primate home ranges and to recognize nonanalog habitats of the past. Our ancestors have run an evolutionary gauntlet of changing climate and vegetation that have spawned many evolutionary innovations, some of them lasting only to the next shift in climate and vegetation, but others of them proven to be of lasting value.

Acknowledgments

I thank my companions of Kenyan expeditions Erick Bestland, Glenn Thackray, Dan Dugas, Jonathan Wynn, Doug Ekart and Hellen Vallianatos. Brenda Benefit, Monte McCrossin and Stephen Gitau helped with transport and accommodation to Maboko, and Meave Leakey was equally hospitable at Kanapoi. Richard and Meave Leakey, Martin Pickford and Christine Kabuye were especially helpful with access to the Kenyan National Museums collections.
References


BAR 1002’00 Orrorin tugenensis femur. Science 305: 1450–1453
Recent advances. Plenum, New York, pp 353–376
Retallack GJ (1992) Middle Miocene fossil plants from Fort Ternan (Kenya) and the evolution of African grasslands. Paleobiology 18: 383–400


14 Quaternary Deposits and Paleosites

Klaus-Dieter Jäger

Abstract

Due to the mineral content of bones and teeth, the majority of fossil hominid remains are represented by these tissues; soft parts of the human body are preserved only very rarely. Whether or not fossils are well preserved depends on the nature of the deposits in which they are enclosed. Numerous methods are now available for chronometric dating of hominid fossils, though none of them is applicable in all situations. However, it is still necessary to situate each hominid fossil within the larger stratigraphic framework. Hominid evolution began well over 4 million years ago, and covered the final part of the Neogene (Upper Tertiary). As a result, ongoing international discussions of stratigraphic boundaries over this span are also significant for the assessment of the hominid evolution. In addition to providing stratigraphic information paleoanthropological sites provide not only insights into the environmental backgrounds of the fossils they yield, but in later periods commonly also into the cultural evolution of mankind and its relatives.

The preservation and suitability for investigation of physical remains left behind by hominids from past centuries or millennia or from periods still longer ago depends mainly on two circumstances:

1. Nature of the fossil remains.
2. Quality and nature of the fossiliferous deposits, as a rule sediments of Pliocene, Pleistocene, or Holocene age.

14.1 Characteristics of fossil remains

Among the preserved remains, hard components from the skeleton are more commonly found, i.e., bones and teeth. Soft parts are more rare.

The preservation of bones and teeth is determined by the content of calcium carbonate in the fossiliferous deposits. In the case of soft parts, their accessibility
to decomposition is determined by the contemporaneous availability of water and atmospheric oxygen.

If one of these two factors is kept to a minimum, the decomposition processes are reduced or stopped.

The access of atmospheric oxygen is reduced in the case of subaquatic sedimentation or sediment conservation (e.g., in inland waters) or in the case of deposits lying below the groundwater table (e.g., peat in boggy terrain). Such conditions characterize the sites of bog bodies. On the other hand, desert areas with minimized water supply provide plenty of atmospheric oxygen access, but the decomposition of organic matter is impeded by the lack of water. In these areas, an essential precondition for a lot of chemical processes is missing. Consequently, under such conditions the decomposition of organic matter is reduced, but frequently the prerequisites for mummification are realized.

More often the preservation of fossil remains is restricted to hard components only. Independent of the local presence of water and atmospheric oxygen, the preservability of bones and teeth is determined by their preservable quality consisting of apatite. This component is characterized by the formula:

\[ 3 \text{Ca}_3(\text{PO}_4)_2 \times \text{CaF}_2. \]

Fossils in limeless deposits are protected from progressive corrosion and destruction since the superficial precipitation is not pure water.

As it passes through the atmosphere prior to precipitation, the rainwater is contaminated by carbon dioxide (CO₂). Consequently, a weak carbonic acid (H₂CO₃) touches every surface, and the lime content in the soil comes in contact with the acid according to the following formula:

\[ \text{CaCO}_3 + \text{H}_2\text{O} + \text{CO}_2 \rightarrow \text{Ca(HCO}_3\text{)}_2 \]

If the subsoil contains a lot of calcium carbonate or a diffuse lime distribution, the first stages of decalcification are concentrated on the lime content of fossiliferous deposits and buried soils, and for a while bones and teeth remain protected against corrosion and dissolution. On the other hand, in limeless deposits, such fossils are the only objects vulnerable to corrosive processes.

### 14.2 Characteristics of fossiliferous deposits

Consequently, bones and teeth in limy deposits are also exposed to corrosive destruction but to a lesser degree. In the case of limy deposits, it is more or less insignificant whether the fossiliferous sediments are loose ones with high porosity (such as sands or loess) or solid rocks (like travertine). Both provide sufficient and favorable conditions for the preservation of human skeletal remains, and this is why finds and sites of paleoanthropological significance are mainly connected
with such deposits. In detail, such conditions are shared mainly by the following sediment types:

1. Fresh-water lime deposits (calcareous deposits from inland waters, as a rule consisting of >90% CaCO₃, frequently modified by diagenesis and especially to be found in travertines)
2. Loess (dust deposits of eolian origin, characterized by lime content, as a rule between 15% and 30% CaCO₃) with dust layers, intercalated by buried soils, frequently with calcareous (and also with decalcified) humus horizons
3. Cave deposits, especially in karstified calcareous mountainous regions, where caves originated from subterranean drainage ways and were occupied by human entry and settlement after drying (which often happened as a result of moving of the original water way)
4. Calcareous debris, especially in mountainous regions, in open-air sites (lime content of the rock detritus being identical to that of the solid rock, perhaps less in the intermediate matter)
5. Calcareous fluvial and deltaic gravels and sands, however preferably containing displaced objects
6. Calcareous lacustrine and beach deposits, as a rule silty and clayey, sometimes laminated

In addition to these site types, those favoring the preservation of soft parts (as in the case of bog bodies) have to be considered, preferably in peat and limnic mud layers.

Moreover, tephras, i.e., pyroclastic deposits traced back to nearby volcanic eruptions—as in the case of Vesuvius during August 79 AD—enable the origination and preservation of human body imprints, which can be replenished by means of gypsum after detection. The method for such a procedure has been applied since the 1860s (since the year 1863 to be precise), first of all by the Italian archeologist Giuseppe Fiorelli who was the longstanding leader of the excavations at the ancient town Pompeii, which was covered by tephra and pumice on the occasion of the aforementioned Vesuvian eruption (Mau 1899). Exceptional finds are the brain endocasts as cited by way of example from the Eemian travertine at Gánovce in Slovakia (Vlček 1958).

**14.3 Possible methods for dating**

In many places, the geological preconditions for fossilized human remains are connected with sites of archeological discoveries proving previous human presence or even settlement. That is why frequently sites of significant paleoanthropological
finds are also highly important in terms of archeology. Not least, such in-site combinations facilitate the dating and the cultural-historical assignation of the respective paleoanthropological observations.

Moreover, the calcareous deposits mentioned above enable not only the preservation of human skeletal remains but also of comparable animal records. The paleozoological investigation of skeletal parts proving previous micro- and macromammals, as well as the determination and examination of fossil shells providing evidence of former mollusks or ostracods, contribute to biostratigraphical datings and paleoecological assessments. In particular, the investigation of mollusk remains is of special validity in Central and Western Europe, since there the precise species determination has been achieved, merely based on conchylia or even their fragments (Ložek 1955, 1964). Moreover, comparable methods of dating are also in preparation in other parts of Eurasia (actually Meng 2007).

However, the investigations of micromammals as well as of mollusks are not restricted to qualitative records. On the contrary, both categories of fossils provide the chance for quantitative analysis and statistical consolidation of the results. Moreover, the examinations permit the reconstruction of faunal assemblages as a basis for the paleoecological interpretation (Rousseau 1990). To such an extent, they contribute to the characterization of the natural environment surrounding previous mankind.

In travertines, frequently the paleozoological record is completed by leaf imprints and incrustations of various plant structures, contributing both to the dating of sites and to their paleoecological characterization.

In sequences of peat and limnic layers of mud or marl, the pollen analysis frequently offers an adequate method for dating. Altogether, paleobotany supports the dating of fossil human remains as well as the reconstruction of the paleoenvironment, especially in the case of travertine and bog sites.

However, all the mentioned paleontological and archeological procedures and observations contribute to relative chronology of sites, whereas the numerical or calendar chronology covering the whole time-span of fossil hominids requires another methodical approach. As a rule, the search for more or less precise dates, both in the geosciences and in archeological disciplines, is focused on so-called absolute chronology. However, already during the early 1980s, the paleontologist Jaeger (1981) claimed that this term is erroneous and—strictly speaking—inadmissible, since this term assumes the existence of absolute time, which is a physical and philosophical nonsense.

Independent of this terminological aspect, this calendar chronology of fossil hominids is mainly based on physical procedures, as summarized ao by Geyh (1980, 1983) and by Wagner (1995, 1998). The preferred methods are provided by
radiometry ($^{14}$C, Uranium series, $^{40}$K/$^{40}$Ar ao) and luminescence procedures (TL, OSL, IR-RF). In addition, the last 10–12 ka is within the reach of the botanical method of dendrochronology, but in practical terms this is restricted to wooden objects. Consequently, paleoanthropological finds can be dated by means of this method in close connection with preserved wood only. In Central Europe, the range of oak chronology provides “a high-resolution time scale for nearby the last 12000 years” (Spurk et al. 1998 p 1114).

However, the preservation of wood as a rule is subject to the same prerequisites as mentioned above with regard to soft parts of human bodies.

### 14.4 Geochronometrical perspective

The choice of applicable procedures suitable for numerical dating and calendar chronology of finds proving the presence of fossil men or early hominids is influenced to a high degree by the characteristics of the fossiliferous deposits. This means different composition of preserved material requires the application of different methods. Consequently, one must accept that there are differences regarding precision and range of dating. As a general rule, the tenet may be assumed that a short range is the inevitable consequence of precise high-time resolution and vice versa. Consequently and independent of the datable material, the older the finds are, the more imprecise is the numerical dating. This aspect is comparable with the optical perspective where the perceptibility of objects decreases with increasing distance. But this is only one aspect of geochronometric perspective.

Another aspect is related to the dated material. Fossil remains of men from historical times and in addition from the last millennia can frequently be dated directly, e.g., by using the radiocarbon method. Fossil remains from the Middle Pleistocene, i.e., finds with an age of a few or some hundreds of thousands of years, are not datable directly, but as a rule the fossiliferous deposits may be subjected to numerical dating, e.g., by means of uranium series in the case of travertine deposits or by application of luminescence procedures in the case of loessic deposits. A larger age characterizes early stages of hominid evolution as evidenced by the occurrence of australopithecines. Such discoveries can be dated using only a few methods, e.g., the potassium–argon method ($^{40}$K/$^{40}$Ar), and as a result merely rocks and loose deposits of volcanic origin can be considered. This means layers of lava rocks and tephras can be dated at sites of fossil hominid remains, where such layers are included in the stratigraphic sequence. Thus, the object of dating is not fossiliferous deposits but distinct layers in the sequence that includes fossiliferous deposits. Ergo, the chronological
investigation is restricted to time-marks in the local stratigraphy of the respective site (cp, e.g., Fitch and Miller 1976; Ullrich 1983: fig. 3).

14.5 Global stratigraphical context

The hominid evolution started more than 4 Ma (cp Johanson and Blake 1996 p 23), but according to previous conclusions of the International Union of Geological Sciences (IUGS) on the occasion of international congress meetings (London 1948; Moscow 1984), this time-span has been subdivided by the boundary between two different geological systems, i.e., the Tertiary and the Quaternary (Aguirre and Pasini 1985; Partridge 1997). The agreement of 1984 was prepared by another international congress meeting held by International Union for Quaternary Research (INQUA) likewise in Moscow in 1982 (Cepek and Jäger 1988). According to this agreement, the chronological position of this boundary has been settled at 1.64 or 1.8 Ma. Both the systems can be characterized by the common feature of long-term significant climatic fluctuations, but nevertheless they are distinguished noticeably by the average level of temperature oscillations according to their mean values, especially in middle and higher latitudes of the globe.

However, recently on behalf of the International Commission on Stratigraphy (ICS), a body of the IUGS with the competence for recommendations related to a world-wide geological timescale, a proposal was presented aiming at the removal of this stratigraphical boundary by extending the preceding system of the Neogene, i.e., the Upper Tertiary, up to the present (Gradstein et al. 2004). Nevertheless, there are serious objections to this proposal (Claque et al. 2004) regarding fundamental environmental changes during the discussed period (Gibbard 2004; Pillans 2004).

Independent of these current discussions, the actual fixation of the debated global stratigraphical boundary at 1.64 or 1.8 Ma seems to be an unfortunate decision (Cepek and Jäger 1988). The chronological base of the younger system or period may be defined more properly at 2.6 Ma (Gelasian stage of the Pliocene series of the Neogene so far: Pillans 2004).

14.6 Environmental dependences of sites

Climatic fluctuations over several hundreds of thousands of years, as mentioned above, determined not only the environmental conditions of previous hominids both in the temperate zones and in polar and subpolar regions but also influenced the human site selection and the chances of fossil preservation.
Varying climate conditions favored different processes of sedimentation and consequently different types of deposits.

This is why during long-term cold periods, the glacial advances of glaciers in subpolar and mountainous regions, was accompanied by periglacial and climatically continental circumstances in the lowlands and hilly uplands of midlatitudes favoring the eolian deposition, mainly of dust. Consequently, hominid finds from such glacial periods have been made rather frequently in loess sequences, as e.g., at Dolní Věstonice in the Czech Republic and in Austrian sites.

On the other hand, many sedimentation processes can take place under warm, frequently under warm and wet, environmental conditions only. The respective deposits could originate either under a constant warm climate, as in the tropics, in lower latitudes or in the warm phases of glacial cycles, the so-called interglacials, in the midlatitudes. This is why the interglacials have favored such deposits as travertine and peat. Travertines may be defined as consolidated fresh-water lime deposits, especially calcareous tufas.

As mentioned above, especially the travertines have provided excellent conditions for the preservation of finds, as well as for environment reconstruction. Moreover, the origination of these deposits required the proximity of springs and waters, due to the dependence of all fresh-water lime deposits on hard water. Owing to the significance of water for human life, sites of travertine formation were also frequently preferred locations of human presence and settlement. Consequently, man has often visited such sites and his remains can be discovered there frequently. Examples are in Central Europe, the well-known interglacial sites at Gáňovce, Tata, Taubach, Stuttgart-Bad Cannstatt, Weimar-Ehringsdorf, Bilzingsleben, or Vértesszöllös. Due to man’s presence and activities at such sites, they are significant not only when seen from a paleoanthropological angle but also as a result of archeological discoveries.

As a rule, interglacial deposits, such as peat, mud, or travertine, contain assemblages of floral and faunal fossils and thus provide chances of quantitative paleontological analyses. Differences related to age are reflected by differences of the assemblage composition recorded by pollen, conchylia, or micromammal bones and teeth. Consequently, suitable methods of quantitative paleontological analysis aiming at biostratigraphical indications and correlations are pollen analysis (peat, mud), paleomalacology and the investigation of micromammal remains (especially in travertines).

Considering differences related to the composition of flora and fauna accompanying fossil hominid remains, a relative chronology of sites may be based on biostratigraphy.

This viewpoint may be exemplified by the application of paleomalacology to famous interglacial travertine sites in Central Europe (Jäger and Ložek 2004),
including significant places of paleoanthropological discoveries (Jäger and Ložek 2005). Sometimes in this way a previous chronological assignation of single sites has to be corrected (e.g., Weimar-Ehringsdorf in Thuringia, Germany: at least Jäger 2001; cp in addition Mania 1993; Steiner 1993).

14.7 Archeological context

The majority of the sites providing fossil human remains, both from early hominids and from modern humans, have enabled archeological observations. To be precise, at many sites bones and teeth prove the previous presence of human beings, but at the same time accompanying archeological finds or man-made modifications of the site tell us about human activities, behavior, and lifestyle. This is why frequently at sites of paleoanthropological significance, archeological discoveries are also made. This is true both with reference to early hominids like Koobi Fora (in the Turkana Basin, Kenya: Coppens et al. 1976) or Hadas (Awash region, Ethiopia: Kimbel et al. 1982) and to later humans.

Since the Middle Paleolithic, a considerable share of the available evidence comes from burials. This dating means that even Neandertals are among the number of specimens recorded in this way (cp summarizing survey by Bosinski 1985 pp 44–52). In this context, among others the famous examples at Le Moustier and La Chapelle aux Saints in France, or the “Old Man” from the Shanidar cave in Iraq, may be mentioned.

During the Upper Paleolithic, the number of burials increased. Among them is a female individual who was portrayed in a contemporaneous ivory carving from the same Pavlovian site at Dolní Věstonice (Moravia, Czech Republic: Klima 1983 pp 83–89). During the Holocene comprising the last 11,600 years (Litt et al. 2001), *Homo sapiens* is represented not only by single burials but rather by multi-individual cemeteries of different age, occasionally comprising hundreds of burials or more (e.g., Hallstatt in Austria, Early Iron Age, mainly 8.–6. century BC—discovered are ca. 1,300 burials until 20. century AD: Kromer 1959; Sacken 1868; later Pauli 1975).

Burials favor the preservation of bones and teeth altogether, frequently even in the original articulation. Apart from burials, frequently the previous attendance of humans at the site of discovery could be proved for other reasons. The preselection of sites of later—or current—evidence could aim at the use of a sleeping site in the case of early hominids, which may be exemplified by the cave system of Swartkrans in Southern Africa (with finds of *Paranthropus* according to Brain 1983).
Later, humans frequently used the sites of record for settlement as well as for activities of everyday life, art, and other cultural purposes. Well-known examples have been recovered in caves as well on open ground.

An outstanding example of cave occupation by previous residents is given by the site 1 at Chou Kou Tien (Zhoukoudian) near the Chinese capital Beijing (Peking), well known as the eponymous site of the so-called Peking Man (Sinanthropus respectively Homo erectus pekinensis: Wei 1988; Huang 1998; Brühl and Laurat 2000 pp 94–100 and pp 9–10). However, evidence of Neandertal men found in caves has proven that these locations were both residential and burial sites. The former are usually distinguished from the latter by the irregular scattering of more or less isolated bones and teeth.

Moreover, the same feature also characterizes residential sites on open ground. However, if calcareous deposits are present in the subsoil to a greater or lesser extent, then such archeological sites also favor the preservation of bones and teeth and more frequently of isolated and often fragmentary objects.

In situ preservation of the original depositional conditions is provided especially at travertine sites (e.g., Vértesszöllös in Hungary: Kretzoi and Dobosi 1990; Bilzingsleben in Germany, recently Mania 1999; Weimar-Ehringsdorf, likewise in Germany: comprising Steiner 1981 and Vlček 1993) and at loess sections (e.g., Dolní Věstonice in Moravia, Czech Republic: Klima 1983; Vlček 1991).

Moreover, fluvial and deltaic as well as—rarely—marine deposits as a rule contain and provide not only isolated and more or less fragmentary but also dislocated parts of the human skeleton. Well-known examples are the famous early mandibles from Dmanisi in Georgia (Dzparidze et al. 1992; summarizing Justus et al. 2000) and from Mauer near Heidelberg in Germany (currently on the local situation: Fezer 1992; Zöller and Stremme 1992; Löscher 1996), as well as the cranium found at Steinheim/Murr in SW Germany (Adam 1988).

On the contrary, not only the preservation of the original articulation of bones and teeth but also of soft parts of the human body is realized frequently in the case of bog bodies (compare among others Hahne 1918; Dieck 1965; Glob 1969; Brothwell 1986; Turner and Scaife 1995; Gebühr 2002).

References


Bosinski G (1985) Der Neandertaler und seine Zeit. Dr Rudolf Habelt Bonn


Hahne H (1918) Die geologische Lagerung der Moorleichen und Moorbrücken, Gebauerschwetschke, Halle/Saale


Jäger KD, Ložek V (2005) On the contribution of palaeomalacology to the stratigraphical placement of Pleistocene travertine sites in Central Europe. Paper in 1st international conference and exhibition on travertine. Pamukkale University, Denizli (Turkey), 2005


Homo erectus heidelbergensis von Mauer (Mannheimer Geschichtsbl NF Beih 1). Thorbecke, Sigmaringen, pp 35–36
Ložek V (1964) Quadrärmollusken der Tscheschoslowakei. Tscheschoslow Akad Wiss, Prag
Mau A (1899) Pompeji, its life and art. Macmillan, New York
Pauli L (1975) Das Gräberfeld vom Salzberg zu Hallstatt – Erforschung – Überlieferung – Auswertbarkeit, Mainz
Sacken Ev (1868) Das Gräberfeld von Hallstatt in Oberösterreich und dessen Altertümer. Wilhelm Braumüller, Wien
15 Zoogeography: Primate and Early Hominin Distribution and Migration Patterns

Alan Turner · Hannah O’Regan

Abstract

Human evolution may be usefully considered as part of the evolution of the larger mammalian fauna of the Tertiary and Quaternary periods. When viewed in this way, the dispersion of the hominins, and the questions of timings and directions, can be examined in the context of movements in other mammalian groups without being treated as special case with a unique set of causes. Earliest migrations by Primates into Africa during the Oligocene and Miocene were accompanied by numerous other taxa that were capable of crossing the closing gap between continents. While the later evolution of the Hominidae may well have taken place in Africa, the dispersion patterns of the Dryopithecine apes during the early Middle Miocene into Eurasia and then back into Africa toward the Late Miocene suggest a rather wider ancestry for humans and the living African apes. Earliest movements across the Gibraltar Straits during the Pliocene and Pleistocene can be ruled out for any terrestrial mammals, and while a Late Pliocene dispersal across the opening Bab-el-Mandeb region of the southern Red Sea may have been possible, such a route appears an unlikely choice for a Pleistocene gateway leaving only the Levantine route across Sinai as a plausible way out of or into Africa.

15.1 Introduction

It is a commonplace observation that species of both plants and animals have patterns of distribution, and that everything is not found everywhere (Cox and Moore 2004). In many cases, such distributions can be explained by the presence of physical or biotic barriers such as mountains, deserts, or water and the absence of suitable foods, while some clearly owe much to modern human interference.
But a species may not have always been where it is found now, while another may formerly have existed in areas from which it is now absent, so that many patterns reflect processes of movement that occurred from a few tens to thousands or millions of years ago. We live on a planet that has been constantly changing, as the continents have shifted and climates have altered, and it is likely that many of the patterns we observe today have been affected by such events.

Change in zoogeography over geological time, and its relationship to tectonic and climatic events, is one of the things that the fossil record can bring to providing a bigger picture of the past, although pitfalls must be overcome in the process. Identification and dating must be accurate, and while presence is clear enough from the fossils, firm evidence for absence in a region may be rather more difficult to distinguish from simple failure to look in the right place or simply to find. But the fossil record is never as good as it will become, and synthesis cannot forever be delayed on the grounds that we do not yet have enough data. As time goes on, and more evidence accumulates from more and more sites, even absence may start to be seen as a real feature of the record and attempts at synthesis of zoogeographic patterns may legitimately be made (Turner and Wood 1993).

Paleoanthropology deals with the evolution of us and our close fossil relatives, the tribe Hominini or hominins, and may be extended, according to the interests of the enquirer, to include the evolution of the Hominioidea (the superfamily that includes lesser apes, great apes, and humans) or even more widely to other Primates. The wider the context, we would argue, the better the understanding of any fossil group is likely to be. In the case of the Hominioidea, one of the most important issues to understand is the question of dispersions, whether of the Primates into Africa in the first place or our closer hominin fossil relatives out of Africa in the later stages of their evolution. The dispersals of the Hominioidea are thus likely to be best understood within the larger context of dispersals in other elements of the mammalian fauna, and our aim in this chapter is therefore to examine that larger context. We begin with a brief discussion of initial movement of the Primates into Africa and then go on to look at the evidence for Miocene to Early Pleistocene distributions within Africa and between Africa and Eurasia.

### 15.2 Early movements of Primates into Africa

Recent molecular analyses of DNA sequences suggest that a small number of mammalian orders (the elephants, hyraxes, tenrecs, aardvarks, elephant shrews, golden moles, and the aquatic manatees or sea cows) together compose a unique
group, the Afrotheria, whose members share a restricted common ancestry (Madsen et al. 2001; Murphy et al. 2001). The fact that the closest (sister) group to these is the South American order Xenarthra (the sloths, anteaters, and armadillos) makes sense in terms of continental separations and offers logical support to the identification of an Afrotheria. A significant absence from the Afrotheria, however, is the Primates, which seem to have evolved in the Northern Hemisphere (Fleagle 1999).

The Afrotheria owe their origins as a distinctive group to the fact that the continent was long isolated after the breakup of the southern landmass of Gondwana. Africa and the southern part of the Arabian Peninsula finally docked with the Eurasian plate in the Early Miocene, around 25–18 Ma, after a northward movement from the breakup estimated at around 14° of latitude (Rögl 1999). In the process of closing with Eurasia, the eastern arm of the Tethys seaway was closed, providing a land connection. At what precise point movements across the shortening gap between Africa and Eurasia became possible for terrestrial animals is unclear, but the Primates are known in some numbers from the Oligocene-age Fayum deposits of Egypt from as early as 33 to perhaps even 40 Ma, and an even earlier appearance around 45 Ma is suggested by material from Algeria (Godinot and Mahboubi 1992). The implication is that some form of island hopping across the shortening gap was possible as tectonic forces buckled the floor of the Tethys and produced a series of small and no doubt short-lived areas of dry land. Following the contact between Africa and Eurasia, a transgression of the Tethys Sea southward left Africa connected to Eurasia via the southernmost part of the Arabian Peninsula across what are now the Bab-el-Mandeb Straits (Tchernov 1992). Much of the later Miocene movement of faunas into and out of Africa therefore probably took place across this region, although our knowledge of the Miocene mammalian fauna of Arabia is currently very poor (Whybrow and Clements 1999).

Those same enormous forces produced by the combined northward movement of Africa and India formed the mountain chains that run from southern Europe to the Himalayas, including the Taurus and Zagros mountains of Turkey and Iran that have combined with the frequently harsh conditions of the Arabian Peninsula to control movements into and out of Africa (Tchernov 1992) (see also later). Toward the end of the Miocene, during the Messinian salinity crisis, tectonic processes in the westernmost region closed the portal with the Atlantic and the Mediterranean began to dry up (Kirksman et al. 1999). During the Early Pliocene the Mediterranean refilled, while the Red Sea widened as the Arabian plate swung away and eventually broke the Bab-el-Mandeb land bridge in the Late Pliocene, part of the eastern African rifting process that continues to the present day. Although some form of land bridge therefore existed between southern Spain and northern
Africa at some point during the Messinian, there is no compelling evidence for one across the Gibraltar Straits since then. Most if not all subsequent mammalian movements between Africa and Eurasia during Plio-Pleistocene times are therefore likely to have been via the Levant and perhaps Arabia (see later).

The better-known Primates of the Fayum were already quite diverse (see Rasmussen Vol. 2, Chapter 3). Numerous genera have been recognised, but their relationships to modern primates have yet to be established. Such diversity points to a well-established presence at that point, but the subsequent record of monkeys is poor until well into the Miocene, when records recommence in the rich deposits of eastern Africa where numerous remains referred to the family Victoriapithecidae occur (Benefit 1999). Primitive apes appear in the fossil record of Africa a little later than the monkeys, and there is a reasonably good record from Late Oligocene to mid-Miocene deposits. One of the oldest specimens comes from Lothidok Hill in northern Kenya, in deposits that may be as early as 27 Ma, while others are found at a range of Early Miocene sites in Uganda and Kenya such as Rusinga Island in what is now Lake Victoria, at Koru and at Songhor in deposits dated to the period 22–17 Ma (Andrews and Humphrey 1999). These animals, belonging to the genera Proconsul, Rangwapithecus, and Nyanzapithecus and varying from large monkey-like up to female gorilla-like in size, are placed in the separate family Proconsulidae and best characterized as “arboreal quadrupeds” (see Ward Vol. 2, Chapter 9).

More advanced apes belonging to the family Hominidae appear in a further radiation in the period 17–12 Ma. They have been referred to the subfamily Dryopithecinae and divided into the tribes Afropithecini, Kenyapithecini, and Dryopithecini, although Andrews (1996) argues for elevation of all to subfamily status, and it is at this stage that we see first evidence for a return movement of Primates out of Africa. Although the Afropithecini occur in Africa and Arabia, the Kenyapithecini are known from Kenya but mostly occur in Turkey and southeastern Europe while the Dryopithecini are European in distribution. In other words, the major known dryopithecine distribution is outside Africa. As we shall show later, this early dispersion from Africa may have implications for our understanding of the later origins of the Homininae, the subfamily containing the African great apes and humans.

The earliest movement of the Primates into Africa may have been “accompanied” by a dispersal of the archaic predator order Creodonta, since members of the family Hyaenodontidae also first appear in Late Eocene deposits there, but further contemporary incursions of other groups are not evident. A clearer pattern emerges at the time of full contact around 20 Ma, when other immigrants from Eurasia included the first perissodactyls in the form of early rhinos, and the bizarre-looking chalicotheres, with their horse-like heads and clawed feet. More
artiodactyls also made their appearances, with the incursion of the giraffoide climactocerids and first antelopes as well as primitive pigs of the genus *Nguruwe*, which must have traversed the continent since they are known from Namibia and Kenya at around 17.5 Ma (Turner and Antón 2004). More advanced cercopithecoid monkeys replaced the earlier Primates, although in situ evolution is hard to distinguish from immigration. True Carnivora entered the continent with the first appearance of cats and the amphicyonid bear-dogs as well as of mustelids (Morales et al. 1998) and, at least in North Africa, of members of the extinct cat-like family Nimravidae, although the creodonts continued to prosper as the dominant meat eaters. In the other direction, the probable dispersion of dryopithecine apes was perhaps preceded by the appearance of anthracotheres in Europe and possibly accompanied by a dispersal of monkeys of the genus *Pliopithecus*, the creodont *Hyainailouros* and the first movement of the proboscideans from Africa (Agustí and Antón 2002). The latter consisted of movements not of elephants proper but of the four-tusked gomphotheres and the deinotheres, the latter marked by a single pair of tusks set in the lower jaw. The subsequent history of some of the various proboscidean genera during the Miocene and Pliocene suggests a good deal of interchange with Eurasia.

On a general note, the structure of the African Early Miocene guild of larger carnivores seems to have divided into flesh eaters among the early cats, nimravids and perhaps smaller creodonts, and bone-crunchers among the amphicyonids and larger creodonts, and by the mid-Miocene the complexity in the guild structure of carnivores was enormous. Such a guild points to an equivalent complexity in the ecological relationships of predators and prey in Africa at this early period. The ungulates do not look particularly well adapted for speed while the predators do not look equipped to chase anything moving particularly fast. Dog-like animals, in the form of some of the smaller hyenas of the genera *Ictitherium*, *Hyaelictitherium*, *Lycyaena*, and *Hyaenictis*, only appear much later in the Miocene (see later) and in the absence of the true dogs of the family Canidae until later still in the Pliocene it is difficult to assess the extent to which pack hunting might have been possible (Turner and Antón 2004). The larger amphicyonids are unlikely to have operated cooperatively and would probably have taken a mixture of carrion and hunted meat. Overall, the zoogeography that we can piece together supports interpretations of the earlier to Middle Miocene vegetation of Africa as generally quite closed (Cerling 1992; Cerling et al. 1997). We might therefore presume that most of the Primates at that time would have found food and refuge among the closed vegetation, and as such have been fairly safe from predation, although some of the earliest cats were small enough to be effective climbers and may have found small monkeys, at least, an attractive target.
15.3 Zoogeographic evidence for the origins of the Hominidae

Africa is usually taken to be the origin point of the human lineage, and so far as the later stages of the Pliocene and Pleistocene are concerned this is generally accepted as true beyond reasonable doubt. Nevertheless some doubts about this matter have been raised in recent years—in particular regarding the origins of *Homo erectus* (White 1995; Dennell 2004)—and it would always be unwise to assume that the fossil record will never surprise us. We return to this point later below. However, as we have seen, the Primates themselves did not originate in Africa, and while emphasizing their incursion during the Oligocene or perhaps even the Middle Eocene may seem like an academic nicety, it is worth stressing that the intermediate period of the mid- and later Miocene witnessed emigrations and perhaps also reimmigrations.

The apes of the subfamilies Dryopithecini and Kenyapithecini that dispersed from Africa underwent a considerable radiation in Eurasia until the end of the Mid-Miocene, around 7–9 Ma (Andrews and Bernor 1999), but in Africa apes are scarce between perhaps 15 Ma and a reappearance at Lothagam at the very end of the Miocene (Andrews and Humphrey 1999; Leakey and Harris 2003). Interpreting such absence in the fossil record is always hazardous, since it may indicate no more than an absence of suitable deposits or inadequate search and recovery. But if it is a real pattern then it is perhaps among those Eurasian advanced hominid apes that we should expect an ancestor for the later great ape and human lineage of the subfamily Homininae. The possibility that the European Dryopithecini make the most plausible candidates has been both proposed (Begun 1993) and questioned (Andrews 1992, 1996; Andrews and Bernor 1999) on several details of taxonomy and systematics, but Solounias et al. (1999) have raised the question again in the context of understanding wider issues of the relationship between faunas of southeastern Europe and Africa. The latter authors point out that many of the savanna-dwelling mammals of Africa may well have originated in what they term the Pikermi Biome, based on the rich Late Miocene Greek locality of Pikermi. They cite somewhat longer necked and thus more advanced giraffes, rhinos of the extant genera *Diceros* and *Ceratotherium*, the false sabretoothed cat *Dinofelis* and the larger bone-smashing hyenas *Belbus beaumonti* and *Adrocuta eximia* as offering primary evidence for such an origin, and it is indeed clear that such animals do make their first appearance in Africa in the latter part of the Miocene. Overall, by around 8 Ma, over the middle part of the Miocene, there is an evidence of considerable incursion from Eurasia generally into Africa if we add to the above list the smaller to midsized and dog-like hyenas of the genera *Protictitherium*, *Ictitherium*, *Hyaenictitherium*, *Lycyaena*, and *Hyaenictis*, the sabretoothed cat
Machairodus, a range of mustelids, and a number of antelopes (Vrba 1995; Werdelin and Turner 1996; Turner and Antón 2004). The impetus for this movement appears to have been a major shift in climate, changing the western European vegetation from subtropical evergreen forest to more deciduous and dry woodland and provoking a turnover in mammalian fauna that Agustí et al. (1999) termed the mid-Vallesian Crisis. By around 9 Ma, the dryopithecine hominids were extinct in Western Europe, although they managed to survive until perhaps 7.5 Ma in Italy and China (Andrews and Bernor 1999). A movement of early hominid apes back into Africa is therefore entirely plausible as part of this larger pattern of dispersion.

15.4 Zoogeography of African Pliocene Hominini

Whether or not the common ancestor of later Hominidae was indeed of Eurasian dryopithecine-like morphology and origin, our African ancestors and relatives changed from being generalized apes to more sophisticated apes with tools over a period of a few million years, only becoming really recognizably human with the earliest appearance of the *H. erectus* lineage at around 1.8 Ma. This transition included shifts to an upright stance and fully bipedal walking and a massive increase in relative and absolute brain size, presumably accompanied by alterations in behavior, social interactions, and intelligence. Details of zoogeography however remain unclear.

The earliest currently known putative hominins are the 7–6 Ma material from Toros Menalla in the Chad Basin referred to *Sahelanthropus tchadensis* (Brunet et al. 2002) and the 6–5.5 Ma *Orrorin tugenensis* from Lukeino in Kenya (Senut et al. 2001). The latter is thought to have been associated with wooded environments, but the Chad Basin paleoenvironment appears to have been more of a mosaic association of lake-side gallery forest, wooded savanna and open grasslands (Vignaud et al. 2002), and the precise preference of the primate is hard to determine.

According to interpretations of the material referred to the later *Ardipithecus ramidus*, first found in 4.4 Ma deposits of the Middle Awash Valley in Ethiopia (White et al. 1994, 1995), an attachment to a woodland habitat may have persisted until close to 4.0 Ma (WoldeGabriel et al. 1994), although it is worth stressing that the authors pointed to similarities to chimpanzee morphology in various features of the remains so that hominin status may yet be rejected. Material also referred to *A. ramidus* from 4.5 to 4.3 Ma sediments to the west of the Awash at As Duma (Semaw et al. 2005) is also said to be associated with moderate rainfall grassland and woodland/grassland, based on paleosols, soil
carbonates, and faunal elements. However, stable carbon isotope values for ungulate dental enamels at As Duma suggest not only browse but also a significant component of C_4 grasses in the diet, and we should beware of the dangers of a small number of early hominin localities misleading us about habitat preferences and true distributions. Further fragmentary material from the Middle Awash Valley dated to even older deposits, around 5.8–5.2 Ma, was referred to *A. ramidus* (Haile-Selassie 2001), but was later given status as a separate subspecies, *A. ramidus kadabba*, on the basis of still more primitive dental characteristics that would presumably place it even closer to a common ancestor of humans and chimps. More recently still, Haile-Selassie et al. (2004) have argued that this form should be referred to as fully separate species, *A. kadabba*, which they consider is very similar to *Orrorin* and *Sahelanthropus*, and thereby infer that Late Miocene diversity was less than would seem to be the case.

We do know that the physical and biotic environment within which the hominins were evolving was itself undergoing significant changes. Rifting, volcanic activity, and uplift were continuing to change the topography of eastern and southern Africa as they had throughout the Miocene (Pritchard 1979; Adams et al. 1996), and as a result of these changes and their interaction with climatic events the vegetation was opening out to provide the distribution and huge mosaic of habitats existing there today. Such physical and biotic changes underlie the distributions of the living African mammal fauna (Grubb 1999; Grubb et al. 1999) and must have had a major bearing on the zoogeography of the past (Turner 1995). Clearly, more open vegetation not only developed during the Pliocene but also became an attractive habitat for predators and in all probability early hominins (Turner and Antón 1999). The development of a stone tool technology can now be traced back in Africa to around 2.4 Ma at Hadar and the Omo Shungura Formation in Ethiopia and the Lokalalei Member at West Turkana (Kibunjia 1994), while the bones of large mammals from Bouri in Ethiopia have damage marks interpreted as hominin-inflicted by means of stone tools (Semaw 2000). What happens in terms of hominin development in the intervening period before such technology was developed, either to assist or to motivate the move to more open terrain, is unclear.

An analysis of the zoogeography of African Pliocene hominins within the larger context of distributions in the rest of the Plio-Pleistocene large-mammal fauna was undertaken by Turner and Wood (1993), prior to the more recent discoveries referred to above but based on a larger body of hominin species and known distributions. The evidence available then, as now, suggested that the genus *Australopithecus* was apparently geographically split, and represented by *A. afarensis* in eastern Africa and *A. africanus* in the south. The genus *Paranthropus*, taken by Turner and Wood to be monophyletic, was considered to be
represented by *P. robustus* and perhaps *P. crassidens* in the south and by *P. aethiopicus* and *P. boisei* in the east. More recent discoveries have extended the range of named hominin taxa in the eastern region but done little to alter that picture of regional distributions, although Bromage and Schrenk (1995) have extended the range of *P. boisei* southward to Malawi. A new and more primitive species, *A. anamensis*, has been identified at East Turkana and Kanapoi in deposits dated to around 4.0 Ma (Leakey et al. 1995) while another, *A. bahrelghazali*, has been recognized at Koro Toro in Chad in deposits of slightly later age (Brunet et al. 1996). A third species, *A. garhi*, has been identified in deposits of the Middle Awash Valley (Asfaw et al. 1999), and a fourth, placed in a new genus as *Kenyanthropus platyops* (Leakey et al. 2001), has been identified from deposits dated to 3.5 Ma at West Turkana.

In the case of the genus *Homo*, assessing the geographic distribution of species is made more difficult by the increasingly evident fact that the taxonomy is more complicated than has previously been assumed. *Homo* is conventionally considered to be evident in Africa back to about 2.5 Ma, first represented in eastern Africa by the species *H. habilis* and *H. rudolfensis*, although the latter has also been identified in Malawi by Bromage and Schrenk (1995). The fact that stone tools appear in the archeological record at about the same time has led to speculations about the relationship between evolutionary change and the development of tool-making abilities. However, it has long been apparent that the earliest taxa assigned to the genus *Homo* are a rather heterogeneous group (Wood 1991, 1992), and arguments that they be removed from *Homo* altogether have recently reemerged, in a scheme leaving *H. erectus sensu lato* as the earliest clear member of the genus (Wood and Collard 1999). This would place the first evidence of our genus almost precisely at the Pliocene–Pleistocene boundary and some time after the earliest appearance of stone tools, and of course it is this seemingly abrupt appearance of *H. erectus* that has led to suggestions of a possible origin outside Africa referred to above (White 1995; Dennell 2004).

Of course, tool making by hominins of other genera is neither improbable nor implausible. Chimpanzees both make and use tools, albeit primitive ones that satisfy any sensible definition of such behavior, and it may be that a variety of evolutionary solutions that included some elements of manufacture and use of technology were developing among the African Pliocene hominins (Turner and Antón 2004). While our own lineage moved toward greater ecological generalization coupled to an emphasis on stone tool technology and an increase in brain size, members of the genus *Paranthropus* appear to have developed larger jaws and teeth in order to cope with their food. At least one hominin species traditionally referred to *Homo*, the east African *H. rudolfensis*, appears to have followed the same path as the paranthropinines between 2.5 and 1.8 Ma with enlarged
teeth but with a relatively large brain as well. If the *H. rudolfensis* material is indeed to be linked to the earlier *K. platyops* and referred to that genus, as suggested by Leakey et al. (2001), we may then very well have evidence of a separate lineage within which brains and teeth both developed. We may therefore identify at least three different evolutionary developments within Pliocene hominins, all of which enjoyed a considerable measure of success.

Whatever the lineages involved, making sense of the zoogeography of Pliocene hominins is difficult. The record presents a complex series of morphologies and proposed taxa, and interpretations of identity, relationship, and adaptations of the various species, leave alone distributions, are impeded by the fragmentary state of much of the material and the fact that several taxa or putative taxa are represented by single specimens or localities. This underscores the value of looking at distributions within the rest of the fauna in order to see whether patterns that may appear to be present in the hominins make sense in terms of larger-scale patterning. The investigation by Turner and Wood (1993) extended to include just such a larger patterning in eastern versus southern African taxa and concluded that there was evidence for a high degree of regional differentiation in some families, particularly the Bovidae, coupled with evidence for significant dispersals in others. Among the Primates, the papionin monkeys appeared to show the most evidence for dispersals, and overall it was apparent that regional isolation was not a matter of rigid demarcation. The implications of this for our understanding of hominin biogeography are that regionally restricted taxa would be a plausible interpretation of the material to hand, but that movements between regions are likely to have occurred and that identified genera such as *Australopithecus* and *Paranthropus* with differing species in each regions are indeed likely to be monophyletic. If monophyly is a correct interpretation, then a localized origin with subsequent dispersals is likely to have been the dominant pattern (Turner and Paterson 1991; Turner 1999a).

The known distribution of Pliocene and perhaps Late Miocene hominin remains now stretches from Chad down through eastern Africa to South Africa, perhaps even from the Atlantic coast of the western Sahel down to the Cape as Brunet et al. (1995) argued, and any reasonable interpretation of that pattern would recognize it as a minimal statement of range. But how much the gaps in between known localities were filled in, or the limits of distribution extended, remains unknown, although we seem to have no evidence for occupation of the northwestern region prior to ca. 1.0 Ma (Raynal et al. 2001). Dennell (2004) has even suggested that if Pliocene hominins were in Chad some 2,500 km west of the Rift Valley by 3.5 mya then why not as far to the north or east by the same period? This, of course, would place them in Arabia or even southwestern Asia.
15.5 Out of Africa

The number, timing, and direction of earliest hominin dispersals from Africa have long been a major point of discussion, and opinions on these topics remain varied and contentious (Rolland 1998; Turner 1999b; Bar-Yosef and Belfer-Cohen 2001; Strauss 2001; Villa 2001; Dennell 2003, 2004). Early Pleistocene assemblages from ‘Ubeidiya in Israel dated to around 1.5 Ma (Belmaker et al. 2002), perhaps Pakistan by 1.8 Ma (Dennell 2004), and human material from Dmanisi in Georgia dated around 1.75 Ma (Gabunia et al. 2001; Vekua et al. 2002) set a minimal date for the original movement. Dispersals to eastern parts of Asia remain more contentious, with the most recent claim based on 1.66 Ma deposits at Majuangou in the Nihewan Basin of northern China with stone implements and what is interpreted as stone-tool processing of animal carcasses (Zhu et al. 2004). The stone tools resemble primitive Oldowan items found in African deposits, and the authors argue for a significant and flourishing early dispersion from Africa, although the rest of the mammalian fauna has no clear African elements. However, the current consensus would seem to favor a shorter chronology after around 0.5 Ma for more extensive and intensive occupation of Europe in particular and Eurasia in general, with a long tail of more sporadic appearances back to and perhaps even beyond the Plio-Pleistocene boundary (Turner 1999b; Roebroeks 2001). But which route, or routes, was used? By land across Sinai and the Levant, across the Bab-el-Mandeb Straits at the south of the Red Sea and then across the Arabian Peninsula proper, or by the Gibraltar Straits?

15.5.1 Gibraltar

Although the evidence available points to some part of the Arabian Peninsula as the most probable route for Plio-Pleistocene hominin dispersions, various claims for possible movements across the Gibraltar Straits have been made most recently by Rolland (1998). He argues for a reduction of the seaway through the Straits to 8 km during glacial maxima, without any increase in surface current and for sweepstake-like movements, especially during OI Stages 12 and 16. Flemming et al. (2003) argue that the Strait itself would not have narrowed significantly during sea-level falls, although they point out that now submerged areas to the west of the Strait would have formed substantial islands that might have provided “stepping stones.” Strauss (2001, p 99) offers a thoughtful analysis of the issue but concludes that the record for human contacts is “at best spotty and ambiguous.” As far as other routes across the Mediterranean are concerned, particularly between Tunisia and Sicily, Flemming et al. (2003) reach no conclusion, and it is not evident that such possible
routes have any strong scientific support. Villa (2001) provides a useful summary of some of the arguments and rejects the idea of such routes.

As summarized elsewhere (O’Regan et al. in press), among the extant and Holocene terrestrial mammals present in North Africa and Iberia, such as wild boar (*Sus scrofa*), red deer (*Cervus elephas*), otter (*Lutra lutra*) and the red fox (*Vulpes vulpes*), there are none that are not also present in the Levant. This suggests that these animals took a circum-Mediterranean route rather than crossing the Gibraltar Straits (Dobson 1998), although of course the possibility of some individual animals swimming across cannot be completely ruled out. Some bat species are found on both sides of the Straits and further eastward in Europe but are not recorded in the Levant or elsewhere in North Africa, which could imply dispersion across the Gibraltar Straits (Dobson and Wright 2000). However, the extent to which recent human interference has played a part remains a concern, and of the 17 terrestrial mammal species inhabiting North Africa today only 4 are considered to have a natural circum-Mediterranean distribution whereas the rest are thought to be recent introductions (Dobson 1998). Clearly, if populations of *H. sapiens* were capable of dispersing over substantial bodies of water to reach Australia in the Late Pleistocene (Bowler et al. 2003), then it is also possible that they may have been moving around the Mediterranean and transporting animals prior to the Holocene, a point stressed by Strauss (2001). However, we believe that the Gibraltar Straits cannot be shown convincingly to have been the scene of natural Pliocene, Pleistocene, or even Holocene movements of terrestrial mammals and that the Arabian Peninsula remains the only established route of a two-way movement between continents.

**15.5.2 Arabia**

In contrast, the possibility of hominin dispersions through the Arabian Peninsula and the Levant, either across the Sinai Peninsula or the Bab-el-Mandeb Straits at the south of the Red Sea, is indicated by the mixed Afro-Eurasian nature of the fauna of the region since the later part of the Pliocene and in particular by the Early Pleistocene deposits at ‘Ubeidiya in Israel (Tchernov 1992; Turner 1999b; Belmaker et al. 2002). Late Pliocene African elements, chiefly bovids and giraffids, are known to the north of the Taurus-Zagros mountain chain that borders the northern boundaries of the Arabian Peninsula at localities such as Kuabebi in the Caucasus and Wolacks in Greece (Sickenberg 1967), the Oltet Valley in Romania (Radulesco and Samson 1990) and Huélagó in southern Spain (Alberdi et al. 2001). The dispersals represented by these occurrences appear to have been part of a larger faunal turnover in Eurasia, what Azzaroli et al. (1988) termed the
Elephant-\textit{Equus} event in Europe, since it was thought to be marked by the first appearance of \textit{Mammuthus} and the true horse, \textit{Equus}, although the first appearance of these genera now seems to have been somewhat earlier (Radulesco and Samson 1990, 2001).

At Dmanisi, the African elements consist of a giraffe, \textit{Paleotragus} sp. (Gabunia et al. 2001), and at Ubeidiya an undetermined giraffid, the bovids \textit{Pelorovis oldowayensis} and an oryx, the hippo \textit{Hippopotamus gorgops}, and perhaps the spotted hyena, \textit{Crocuta crocuta} (Tchernov 1992). The Acheulean industry at Ubeidiya is unknown at Dmanisi and appears to have developed in Africa at a slightly later date. Other Early Pleistocene localities in Europe have relatively few African species. Much has been made of the appearance of the cercopithcoid \textit{Theropithecus cf. T. oswaldi} at Cueva Victoria in southeastern Spain (Gibert et al. 1995) and Piro Nord in Italy (Rook et al. 2004) and of the possibly African machairodont cat, \textit{Megantereon whitei}, at Venta Micena and in the fauna at Dmanisi and at the Greek locality of Apollonia (Martínez-Navarro and Palmqvist 1995, 1996). However, while \textit{Theropithecus} may indicate African links, the genus \textit{Megantereon} was a well-established member of the Afro-Eurasian carnivore fauna after around 3.0 Myr (Turner 1987) so that claims for its significance in the question of dispersals require further investigation. Of rather more obvious importance is the presence of \textit{Hippopotamus}, which makes an early appearance at the Pliocene locality of Valea Graunceanului in Romania (Bolomey, 1965), and at several the Lower Pleistocene sites such as ‘Ubeidiya and Venta Micena (O’Regan et al. in press).

Overall, we conclude that the suggestions of possible hominin dispersions into Eurasia during the later Pliocene that appear in the literature from time to time (Bonifay and Vandermeersch 1991; Boitel et al. 1996), while unsupported by critical assessments of the evidence within Eurasia, cannot be dismissed \textit{a priori} as impossible or even unlikely. These conclusions parallel some of those reached by Mithen and Reed (2002) in their computer simulation of dispersals and stressed elsewhere (Dennell 1998, 2004). If recent arguments summarized earlier about the status of later Pliocene species referred to the genus \textit{Homo} are correct, then clearly the earliest hominin to have moved out of Africa would not have been a member of our own genus. However, whether such a view offers support for an extra-African origin for \textit{H. erectus} remains unclear to us.

\section*{15.5.3 Precise routes}

If Arabia is indeed the only plausible route out of Africa for terrestrial mammals during the Plio-Pleistocene, it remains difficult to judge the relative contributions
of the Sinai versus the Bab-el-Mandeb as gateways. In a recent paper, Petraglia and Alsharekh (2004) outline some of the problems that beset understanding the Middle Paleolithic or later Middle Pleistocene hominin occupation of Arabia, and the relative paucity of evidence for Lower Paleolithic occupation at the small number of known localities. However, as Petraglia (2003) also showed, both Oldowan and Acheulean assemblages are known, and the eastern side of the Bab-el-Mandeb Straits in particular appears to have been occupied by hominins with this technology. Unfortunately, the absence of good chronological control remains a major obstacle to assessing the pattern of lithic assemblage distribution, while research focus and the difficulty of fieldwork add to problems in overcoming our lack of knowledge of the area. But while these obstacles to understanding are clearly real, it remains hard to believe that much of the Peninsula was exploitable for much of the time by latest Pliocene or Early Pleistocene hominins.

As Petraglia and Alsharekh (2004) show, while movement across Sinai offers the possibility of movement along the Levant and then perhaps south into Northern Arabia either along the eastern Red Sea Coast or inland behind the highlands of the Hejaz Asir, movement across Bab-el-Mandeb confronts any dispersing population directly with the highlands. These would tend to restrict movement to the coastal strips, north along the eastern Red Sea Coast or east along the Arabian Sea Coast. While annual rainfall today in the Hejaz Asir or in the Oman Mountains at the easternmost corner of the Peninsula can reach well over 100 mm, much of the southeastern portion, the Rub Al Khali or empty quarter, may have no more than 50 mm with temperatures that exceed 50°C (Glennie and Singhvi 2002). The fact that Lower and Middle Paleolithic occupation did occur means of course that conditions were not always unfavorable, and as Glennie and Singhvi show, the presence of substantial alluvial fans suggests that some earlier interglacials may have been more humid than today, although again the lack of chronological control hampers interpretation. We also know little of the Plio-Pleistocene fauna of Arabia, which is even more sparsely represented than that of the Miocene. The sole exceptions are the small assemblages from An Nefud in northern Saudi Arabia, thought to be of Early Pleistocene age (Thomas et al. 1998) and which, with spotted hyena, hexaprotodont hippo, horse, elephant (cf. *Elephas recki*), several bovids including a species of *Pelorovis* as well as crocodile and fish, are of distinctly African stamp. The range of species implies good grassland and standing water in the vicinity, an interpretation supported by isotopic analyses of herbivore teeth.

However, as Glennie and Singhvi also point out, increased aridity is indicated during glacial periods, beyond even that seen today despite the fact that temperatures may have averaged somewhat lower, and this factor presumably played a large part in determining the extent to which mammals, including early
hominins, could maintain any occupation throughout the Pleistocene. Glacial periods with their massive falls in sea level are of course precisely the point at which the Bab-el-Mandeb crossing is likely to have been at its most obvious and navigable to early hominins (Rohling et al. 1998; Cachel and Harris 1998), so that the easiest and most attractive access by that route is likely to have been at a point when conditions in southern Arabia, and for that matter on the corresponding coastal area of Africa, are likely to have been least appealing. Taken overall, the Bab-el-Mandeb Straits do not seem likely to have offered a likely gateway out of Africa for terrestrial mammals during the Lower and Middle Pleistocene, suggesting that movement across Sinai and then northward along the Levant, southward into Arabia, or eastward and beyond is the most plausible route. However, movement across the Bab-el-Mandeb during the later Pliocene, before the Straits had fully formed, and thus before the crossing into Arabia was dependent on sea-level fall, may have been an entirely different matter, as previously pointed out (Turner 1999b). The importance of the Afar region of Africa to the south and west of Bab-el-Mandeb as an area of attractive resources for mammals, including hominins, following rivers into the developing depression as rifting progressed from Miocene times onward was highlighted by Kalb (1995, p 366), who stressed “the step-by-step process of animal migrations into and dispersal across intercontinental areas prior to complete plate separation.”

15.6 Conclusions

It seems clear to us that the wider context of movements between Africa and Eurasia throws useful light on the patterns of dispersion within Primates in general and Hominoidea in particular so that we can examine changes in primate distribution without having to treat them as a special case. In the earliest stages of the Oligocene and Early Miocene, we see that movements into Africa are accompanied by a range of other taxa and that the nature of the earliest guild of mammalian predators would perhaps have posed little threat to a largely arboreal range of Primates there. If dryopithecine apes first moved back into Europe during the Mid-Miocene and then back again into Africa toward the end of that epoch, then they did so as part of a much wider dispersal across a range of mammalian orders.

Our current knowledge of Pliocene hominins is at best incomplete, and while recent discoveries have extended the range of named species, they have done little to clarify the likely relationships between those taxa or the true nature of distributions. With the greater number of putative taxa now available, we suspect that it will require some considerable time and quite a few more discoveries
before real sense can be made of the patterns. However, as far as movement out of Africa by hominins is concerned, we believe that a number of points can be made. The Gibraltar Straits are unlikely to have been the site of any Pliocene or Pleistocene gateway for terrestrial mammals; and while movement across the Bab-el-Mandeb region before the Red Sea opened fully toward the end of the Pliocene may have been possible, dispersal across the Straits once the glacial and interglacial cycle got underway seems to us unlikely. Sinai and the Levant, the scene of a two-way faunal movement between continents during the later Pliocene, remains the only established route. Moreover, the diversity of elements using that gateway points to the relatively hospitable nature of the area in the later Pliocene, so that suggestions of possible hominin dispersions into Eurasia during the later Pliocene are not inherently implausible. However, how long any dispersing populations were able to maintain their extended range is entirely another question. We suspect that early hominin movements into Arabia and the Levant were sporadic and probably tenuous and that subsequent movements out into Eurasia proper were probably even more so.

References

Azzaroli A, De Giuli C, Ficcarelli G, Torre D (1988) Late Pliocene to Early Mid-Pleistocene mammals in Eurasia: faunal succession and
dispersal events. Palaeogeogr Palaeoclimatol Palaeoecol 66: 77–100


earliest hominin movements. Cour Forsch Senckenberg
Turner A (1987) Megantereon cultridens from Plio-Pleistocene age deposits in Africa and Eurasia, with comments on dispersal and
the possibility of a New World origin (Mammalia, Felidae, Machairodontinae). J Paleontol 61(6): 1256–1268


Villa P (2001) Early Italy and the colonization of Western Europe. Quat Int 75: 113–130


Abstract
The history of life on earth, from the earliest microscopic cells to the modern world populated by the rich variety of animals, plants, fungi, and microbes, is more than 3500 Myr long. Documenting the diversity patterns through the Proterozoic and Phanerozoic has been a major task in the past decades and is fraught with many methodological problems. The emergent picture is one of a very irregular increase in diversity. The most significant episodes of diversification occurred during the Cambrian–Ordovician and throughout the Mesozoic–Cenozoic. In the Phanerozoic alone, 5 major and more than 15 smaller mass extinctions disrupted the diversification of life and sometimes drastically altered the way of evolution. There was no common cause for these events, but all were the consequence of large-scale environmental perturbations. There is growing concern that we are currently entering a “Sixth” major extinction, caused by human impact on nature.

16.1 Introduction
This chapter reviews the history of diversity, the “ups and downs” of life, during the past 3500 Myr. There are many reasons why the history of life’s diversity on earth is an important avenue of paleontologic research. (1) Such studies can give us insight into the relative importance of various mechanisms of evolution. (2) The analysis of diversification and extinction can help to clarify the respective roles of biotic (intrinsic) versus abiotic (extrinsic) factors. (3) The importance of regional versus global patterns/mechanisms can be investigated. (4) The analysis of extinctions might enable us to see if there are common patterns. (5) The possibility that Homo sapiens is causing yet another mass extinction fosters interest in other such events in the geological past.

One emergent theme when documenting patterns of diversification and extinction is that the major disruptions/discontinuities were not the consequence of some lineages’ “racial senility” or “genetic exhaustion” but rather caused by
large-scale environmental perturbations. The world as we know it today with its current climatic, atmospheric conditions, and present-day biosphere is not a good actualistic example for the remote past, where sea levels and mean annual temperatures were vastly fluctuating, continental plates had entirely different positions, and CO\textsubscript{2} and O\textsubscript{2} levels in the atmosphere showed secular changes.

This chapter is organized as follows:

- First an overview of major events in the history of life is given, together with an overview of changes in abiotic conditions. Here some emphasis is placed on the Precambrian.
- The patterns of diversification of Phanerozoic life is reviewed, most extensively for marine animals with preservable hard parts.
- Methodological problems in documenting diversity (as a measure of taxonomic richness) through time are discussed.
- In a special subchapter, some of the most important radiations and extinctions are treated in stratigraphical order.

16.2 **A short history of life on earth**

The earth is roughly 4600 Ma old ([Figure 16.1](#)), and the oldest rocks currently recognized (Acasta gneiss from the Northwest Territories, Canada) are dated as 4030 Ma (Nelson 2004). The earliest Eon of earth history, the Hadean, thus left no directly observable documents. It is nowadays assumed that the initially high temperatures following the accretion of the earth dropped around 4000 Ma below 100°C. During the earliest Archean, the surface temperature was probably quite low (faint early sun). By that time, liquid water was present on the earth’s surface, brought to the earth by comets (McClendon 1999). Yet the heavy bombardement with bodies exceeding 250 km lasted until 4200–3800 Ma. This must have led to repeated boiling of the oceans and the vaporization of the water (Nisbet and Sleep 2001, 2003).

Contrary to earlier beliefs (e.g., those assumed in the famous Miller experiment), most researchers today think that the Hadean and early U.S atmosphere was only mildly reducing, with mainly CO\textsubscript{2} and N\textsubscript{2}, but also smaller amounts of CH\textsubscript{4}, NH\textsubscript{3}, H\textsubscript{2} present (McClendon 1999; Raven and Skene 2003). Free oxygen builded perhaps through photodissociation of water vapor in the upper atmosphere but quickly reacted with Fe\textsuperscript{2+} and other unoxidized compounds. The early atmosphere was, therefore, devoid of free oxygen (McClendon 1999; Miller and Lazcano 2002).
Figure 16.1
Major events in earth history over the past 4600 Ma. Modified after various sources.
Under these conditions life originated (for reviews see Oró et al. 1990; Brack 1998; McClendon 1999; Fenchel 2002; Taylor 2005). Organic compounds could form in the atmosphere through UV-photolysis, electrical discharges, and major impact shocks (McClendon 1999; Lazcano 2001), and a major source of organic molecules was probably also comets and interplanetary dust (Lazcano 2001; Miller and Lazcano 2002), but early life certainly evolved in the sea (Raven and Skene 2003), and the first cells were probably heterotrophs (Lazcano and Miller 1996; but see Huber and Wächtershäuser 1997; Wächtershäuser 2000). As possible sites where cellular life evolved from prebiotic precursors, continental thermal springs, volcanic vents, warm hypersaline lagoons (Darwin’s “warm little pond”), and deep submarine vents are the most likely candidates (Schopf 1999). In this respect, it is noteworthy that the earliest branches of both the Archea and Bacteria are thermophilic (Pace 1997). Yet an extraterrestrial origin of cellular life cannot totally be excluded (Horneck 2003).

Perhaps life originated sequentially several times but was always exterminated by the heavy meteorite bombardment (“impact frustration”) until life took hold around 3900 Ma (Schopf 2002). The “Universal Ancestor” might have been a diverse community of cells that experienced extensive horizontal gene transfer (Woese 1998, 2002). Organismal lineages established themselves only with the subsequent splitting into the three domains “Bacteria,” “Archea,” and “Eukarya” (Woese 1998). The tree of life is thus at its base rather a web (Doolittle 1999) or even a ring of life (Rivera and Lake 2004).

The fossil record of the Archean is notoriously sparse. The presumed oldest cellular microfossils come from 3500 to 3450 Ma sediments in Western Australia (Schopf 1992a, 2004; but see Brasier et al. 2004). Almost as old (3400–3200 Ma) are microfossils in the metasediments of the Onverwacht and Fig tree groups (South Africa). From even older (3800 Ma) metamorphosed sedimentary rocks from Greenland (Isua Greenstone belt), putative microfossils (Isuasphaera) were described (Pflug 1978). Yet they are probably of inorganic origin (Bridgwater et al. 1981; Appel et al. 2003).

According to most authors (Mooers and Redfield 1996; Sheridan et al. 2003) the split between bacteria and archea–eukaryota must have occurred more than 3500 Ma (but see Doolittle et al. 1996). The Proterozoic fossil record (Hofmann and Schopf 1983; Mendelson and Schopf 1992; Schopf 1992a, b) documents an increasing diversification and complexity of cells but the evolution of metabolic pathways cannot be deduced from the morphology of the microfossils. Yet by 2700 Ma, oxygenic photosynthesis, methanogenesis, and methylotrophy had probably developed, perhaps also sulfate reduction and nitrogen fixation (Buick 2001).

In contrast to cellular fossils, the record of stromatolites is quite good. These are biosedimentary structures built by microbial mats which trap sediment
particles (Riding 1991; Walter 2001). The oldest stromatolites date from 3500 to 3200 Ma sediments in Australia and South Africa, but their fossil record remains spotty until the Neoarchean (Walter 1983, 2001; Grotzinger and Knoll 1999). They diversified considerably in the Proterozoic, and in the Mesoproterozoic, a large number of different types had developed, growing sometimes to impressive sizes (Awramik and Sprinkle 1999; Walter 2001). Stromatolites declined both in diversity and abundance in the terminal Proterozoic, and in the Phanerozoic, they remained largely restricted to marginal marine environments. This decline was traditionally seen as a consequence of the rise of the metazoans, which consumed the microbial mats, but it might also have abiotic reasons (Awramik and Sprinkle 1999).

2450–2200 Ma, the earth witnessed the first well-documented ice age, the Huronian glaciation. As a consequence of insufficient time resolution, the exact duration of glacial intervals during this Paleoproterozoic time period is unknown (Young 2004). The onset and the termination seem to have been gradual (Young 2004). Perhaps an increase in atmospheric oxygen (see later) was the cause of the Huronian glaciation (Kopp et al. 2005). If methane was a major contributor to a greenhouse effect prior to that time, it would have been oxidized in a more oxygen-rich atmosphere, and the removal of this greenhouse gas could well lead to global cooling and the onset of an ice age.

At some time between 2500 and 1900 Ma, the atmosphere changed rather abruptly from reducing ($pO_2 < 1\%$ present atmospheric level PAL) to oxidizing ($pO_2 > 15\%$ PAL; Figure 16.1), largely as a result of the activities of oxygenic photoautotrophs (Schopf 1992a; Holland 1994; for alternative scenarios of atmosphere evolution see Ohmoto 2004). Until about 2200 Ma, free oxygen was constantly removed by oxidation of weathered reduced minerals (Nisbeth and Sleep 2001; Lenton 2003). Detrital pyrite and uranite are common until 2200 Ma and indicate very low levels of free oxygen; otherwise these minerals would have been oxidized (Schopf 1992a; Holland 1994). Likewise, the genesis of banded iron formations (BIFs), which are absent in rocks younger than 1700 Ma, requires reducing conditions in the deeper parts of the seas (Simonson 2003).

The origin of the eukaryotes can be placed at some time before 2100 Ma. The oldest, currently recognized eukaryote is the 2100 Ma multicellular “algae” Grypania (Han and Runnegar 1992). Perhaps the origin of eukaryotes was linked to the increased oxygen content of the atmosphere and shallow waters where a nucleus and its protective membrane would be advantageous (Dyer and Obar 1994). Subsequent evolution of the eukaryotes took place by serial endosymbiosis (Margulis 1981) in which engulfed $\alpha$-purple-bacteria became mitochondria and cyanobacteria became plastids (Dyer and Obar 1994; Pace 1997). In part, even secondary endosymbiosis must have occurred in which photosynthetic eukaryotes were engulfed by nonphotosynthetic eukaryotes (Woese et al. 1990).
Little is known about the early evolution of the eukaryotes. Apart from the acritarchs that are a heterogeneous assemblage of planktonic, unicellular eukaryotes known from the Late Paleoproterozoic and extending into the Phanerozoic (Martin 1993; Knoll 1994), other fossils of the eukaryotic clade include various carbonaceous films that are not easy to interpret (Hofmann 1994). Yet new molecular dates, which are in reasonable agreement with paleontological findings, give some clues as to when new groups originated. The divergence between protists and crown-group plants may be as young as 1100 Ma, the origin of the fungi is placed at 1000 Ma, the split between choanoflagellates and eumetazoans occurred at 900 Ma, and the first bilaterians might have made their appearance at 700 Ma (Douzery et al. 2004; see also Peterson et al. 2004; Peterson and Butterfield 2005). Yet these dates are not universally accepted, and previous molecular-clock estimates have yielded considerably older dates (see review in Erwin and Davidson 2002).

Between 800 and 600 Ma, the earth witnessed several large glaciations. The last two episodes are well dated. The Sturtian glaciation occurred around 710 Ma, and the Marinoan (Varanger) glaciation ended at 630–600 Ma (Allen and Hoffman 2005). Some, perhaps most, continental land masses had a near-equatorial position at that time. Glaciers might have reached equatorial regions and perhaps extended down to the sea level. According to the most dramatic scenario (Hoffman et al. 1998; Hoffman and Schrag 2002; but see Chandler and Sohl 2000; Poulsen 2003), temperatures initially dropped due to some unknown mechanism, but as soon as glaciers reached a critical extension in low latitudes, enough solar energy was reflected back into space that ice sheets could grow at an ever increasing rate. In this “runaway albedo” model, not only all the land masses became ice covered, but the surface of the oceans were also globally frozen (“snowball earth”; Kirschvink 1992; Hoffman et al. 1998; Hoffman and Schrag 2002). The seas became anoxic and BIFs could accumulate again (Hoffman et al. 1998). Eventually, enough volcanic CO₂ accumulated, and the glaciations ended abruptly. Later Neoproterozoic glaciations are documented, but these were not global in their extent (Knoll et al. 2004).

Shortly after the last of these major Neoproterozoic glaciations, the first metazoans enter the fossil record. Enigmatic soft-bodied fossils 610–542 Myr old are known from localities around the world and named Ediacara assemblages (after the Ediacara Hills in South Australia). Although the nature of the flattened, segmented, or quilted Ediacara organisms is still disputed (mainly cnidarians and annelids; Glaessner 1983, 1984; Jenkins 1992; or organisms not related to any of the extant animal phyla; Seilacher 1989, 1992; Buss and Seilacher 1994), they are accompanied by traces produced by bilaterian metazoans (Knoll and Carroll 1999; Valentine et al. 1999; Martin et al. 2000). Yet tracemakers were small and
rare and did not disrupt the sediment. This only changed near the Precambrian–Cambrian boundary when bioturbation markedly increased and the sediment became unstable. This probably led to the extinction of the immobile Ediacara organisms (Seilacher 1999).

Within a short time after the beginning of the Cambrian at 542 Ma, the most remarkable episode in the history of life started, the so-called “Cambrian explosion.” Within only 20 Myr, all animal phyla with preservable hard parts (with the exception of the bryozoans) appeared (Knoll and Carroll 1999; Erwin 2001a; Valentine 2002, 2004). With the beginning of the Cambrian and continuing throughout the Phanerozoic, we have a very reliable fossil record especially for the marine invertebrates with easily preservable hardparts. Fossiliferous localities that show exceptional preservation are interspersed in the stratigraphical record and provide us with information on the soft-bodied fauna (Bottjer et al. 2002). Information on the Phanerozoic history of life can be found in most textbooks on paleontology and historical geology (Cowen 2005; Stanley 2005), and extensive treatment is beyond the scope of that chapter. Details on selected episodes of radiation and extinction are given in the last part of this chapter.

During the Cambrian, life was exclusively marine and dominated by trilobites and a variety of other arthropods. Because the trilobites were highly diverse, are easily determined, and show a high species turnover, trilobites are the most important index fossils for this period. Brachiopods were small and belonged mostly to the inarticulate groups. Among the mollusks, hyoliths and monoplacophorans were the most conspicuous ones, but in the Late Cambrian, the first small nautiloids appeared which marks the beginning of a highly successful group of marine predators. Cambrian echinoderms belonged mostly to groups that were immobile and became extinct during the early Paleozoic. Already during the Early Cambrian, the first true reefs built by the spongelike archeocyathans developed but archeocyathans became extinct at the end of the Early Cambrian, leaving a reef gap until the Middle Ordovician. The first chordates and agnathan fishes appear to have been rare with the exception of the conodonts.

The extinction events in the Late Cambrian affected most severely the trilobites and several echinoderm groups. Ordovician and Silurian seas became to be dominated by articulated brachiopods and stalked echinoderms (crinoids and blastoids). Although the first deep burrows appeared at that time, life was still mainly epibenthic. Large reefs dominated by tabulate and rugose corals and stromatoporoids developed, and bryozoans became an important component of marine hard bottoms. Among the planktonic organisms, the graptolites diversified considerably and have proven to be the most valuable index fossils for the Ordovician and Silurian. Large predators developed among the nektonic nautiloids and among the eurypterids. Several groups invaded freshwater
environments, among them the arthropods and various fish groups. Colonialization of the land started in the Silurian, first by plants (although algal crusts and fungi may have been present earlier), then by mites, arachnids, millipedes, and scorpions. Among the vertebrates, the appearance of the first jawed fishes in the Late Silurian was a major innovation.

During the Devonian and Carboniferous, land plants diversified in an explosive manner leading to the first true forests, and the first seed plants had developed by the Carboniferous. The first wingless insects appeared in the Early Devonian, but it was not until the Carboniferous that the first winged insects conquered the air. In the vertebrates, huge marine predators developed among the placoderms. The Devonian, also called the “age of fishes,” saw the first appearance of the chondrichthyan and actinopterygians while the agnathans considerably diversified. The first tetrapods appeared during the Late Devonian, and the amniote egg evolved in the Late Carboniferous. Among the marine benthic organisms, tabulate-rugose-stromatoporoid reefs attained a new climax. Articulate brachiopods and stalked crinoids still dominated most seafloors. Among the nektonic organisms, the evolution of the ammonoids was a major innovation during the Early Devonian. This group provided the most important index fossils for the Devonian to Cretaceous periods. In the latest Paleozoic, life in the seas did not radically change, but the absence of the heavily armored fishes is notable and among the cephalopods, the stoutly shelled nautiloids also declined while the ammonoids flourished. Reef building was confined to smaller constructs after the Late-Devonian mass extinction. On land, the mammal-like reptiles were the dominant herbivores and carnivores.

After the end-Permian mass extinction, life in the seas and on land dramatically changed. Seafloors were no longer dominated by epifaunal brachiopods and crinoids but by gastropods and burrowing bivalves. Reef production came to a halt during the Early Triassic, and it was not until the Middle Triassic that scleractinian coral reefs became established. This type of reef building would dominate throughout the rest of the Mesozoic and Cenozoic. Only during part of the Cretaceous were the corals replaced as principal reef builders by a group of aberrant bivalves, the rudists. The open waters were dominated by ammonoids and actinopterygian fishes, and during the Mesozoic, three groups of planktonic organisms which play an eminent role in the biogeochemical cycles made their appearance: the planktonic foraminifers, the coccoliths, and the diatoms. The largest creatures of the seas and the top predators were the marine reptiles: nothosaurs, plesiosaurs, ichthyosaurs, marine crocodiles, turtles, and mosasaurs.

On land, the conifers, cycads, and ginkgos flourished and dominated the forests until the late Early Cretaceous, when they became increasingly replaced by the flowering plants (angiosperms). Among the amphibians, the last stegocephalians
died out while the first frogs and salamanders appeared. After the end-Permian mass extinction, the mammal-like reptiles had become marginal players, and the archosaurs, most notably the dinosaurs, became the rulers of the earth. The pterosaurs were the first vertebrates to conquer the air by the Late Triassic, and in the Late Jurassic, the first birds evolved. Mammals arose already during the Late Triassic but remained mostly small and peripheral throughout the Mesozoic.

The mass extinction at the Cretaceous–Tertiary boundary, although much less severe than the extinction at the end of the Permian, nevertheless severely altered the structure of both marine and terrestrial ecosystems. In the seas, the shelled cephalopods became entirely extinct with the exception of a small group of nautilids, and from the marine reptiles only the turtles survived. The vacant ecospaces were filled up by actinopterygian fishes and mammals. The mammals also played a central role in the restructuring of the terrestrial communities. During the remarkable radiation in the Paleogene, mammals occupied almost all available niches on land, invaded the seas (whales, pinnipeds, sea cows), and conquered the air (bats). Birds equally radiated considerably, and in some regions large, flightless birds even became the top predators. After the Eocene–Oligocene climatic revolution, the flora and fauna had an essentially modern organization, and grasses as the most important terrestrial producers became widespread during the Miocene. The Miocene radiation of the apes ultimately led to the development of our species, *Homo sapiens*.

### 16.2.1 Directionality versus contingency

During the Precambrian, eukaryotes evolved from prokaryotes, and in the Cambrian, most animal phyla evolved. Ever more environments were colonialized during the Phanerozoic, including the land and the air. For the evolution of the vertebrates, the long-portrayed succession of the “age of fishes,” “age of reptiles,” and “age of mammals” is of course a gross oversimplification. Osteichthyans and chondrichthians were never as diverse as today! Yet although measuring morphological complexity is not straightforward, there was clearly a trend toward increasing complexity in the evolution of life from bacteria to today’s biota (Carroll 2001).

There are two fundamentally different views on this complexity increase. According to one view, the rise and ultimately dominance of complex life forms were an inevitable outcome of natural selection whereby more complex organisms outcompete the more primitive ones. It also inevitably culminates with the development of intelligent animals. Such a view is championed, e.g., by Conway-Morris (1998, 2003) and is reminiscent of an Aristotelian “chain of being.”
A radically different view of the evolution of life does not see complex mammals as superior to bacteria. The increase in diversity is not neglected but explained in a different way. Bacterial cells cannot evolve toward ever smaller sizes, unicellular eukaryotes cannot evolve toward zero cells but only toward multicellularity, and marine organisms cannot evolve to colonialize a nonexisting environment, but they can evolve land-dwelling species, and among animals also flying taxa. The evolution toward greater complexity is, thus, simply a move away from an “impermeable left wall” and an increase in variance, but in a sense we still live in a bacteria-dominated world (Gould 1996).

Perhaps the truth lies somewhere in between. There is no necessity to evolve toward intelligence, but ultimately this was made possible through successive evolutionary steps after the increasing variance crossed several thresholds. In such a view, life on earth followed a megatrajectory along the following sequence (Knoll and Bambach 2000): prokaryote diversification (including metabolic pathways, but inability for sexual reproduction), early eukaryote diversification (first consumers, multicellularity, increased size, sex), aquatic multicellularity (large size, packaging of biomass, fast movement, complex food chains), invasion of the land (huge biomass of producers, adaptation to widely fluctuating environments), and ultimately intelligence (which perhaps could also have evolved in the water). Nevertheless, if “life’s tape would be replayed,” the outcome would certainly be totally different (Gould 2001).

16.3 Diversity patterns: the broad picture

16.3.1 The current consensus

The 1970s and 1980s witnessed a veritable boom of publications which documented the diversity patterns of life during earth history, the pace of large-scale evolution. Yet the achievements of the late “Jack” Sepkoski clearly stand above those of all others at that time. His 1981 curve (Figure 16.2) of Phanerozoic diversity patterns of marine families (and subsequent refinements down to the genus level; Sepkoski 1996) is one of the most often reproduced figures in the paleontological literature (equaling perhaps the Berlin Archaeopteryx specimen; Benton 2001), and this benchmark is an ideal starting point for the discussion of life’s ups and downs during the last 542 Myr.

Assembling literature-based compilations of the worldwide temporal distribution of marine families (Sepkoski 1982, 1992) and genera (Sepkoski 2002), the number of described taxa and hence diversity for every successive stratigraphical interval (usually stages) could be determined. In addition, the
times of origination (first occurrence) and extinction (last occurrence) of these taxa could be deduced.

By means of a factorial analysis, Sepkoski (1981) could show that the marine families grouped into three rather well-delineated entities each showing a common pattern of origination, diversification, and decline. These three “Evolutionary Faunas” were called “Cambrian Fauna,” “Paleozoic Fauna,” and “Modern Fauna” (Sepkoski 1981). A division into finer intervals (nine “Ecological Evolutionary Units”) was suggested by Boucot (1983) and further refined by Sheehan (1996, 2001a); yet these intervals mainly portray ecological communities and not evolutionary units.

Diversity at the family level shows a sharp rise at the beginning of the Cambrian, which continues well into the Middle Cambrian. In the Middle Cambrian, a plateau was reached and several drops in diversity occurred. The main representatives of this Cambrian fauna were trilobites, inarticulate brachiopods, hyolithids, monoplacophorans, archaeocyathids, and eocrinoids (Sepkoski 1981, 1984). During the Early Ordovician, diversity increased again to reach a level about three times that of the Cambrian. Responsible for that fast and unprecedented increase in diversity were members of the Paleozoic fauna, mainly articulate brachiopods, crinoids, rugose and tabulate corals, ostracods, cephalopods,
bryozoans, asteroids, ophiuroids, and graptolites. The diversity reached by the Ordovician radiation remained almost constant through the rest of the Paleozoic, interrupted only by punctuations of extinctions and subsequent recoveries.

The Late Permian witnessed a dramatic extinction pulse, and the members of the Paleozoic Fauna lost their dominance. The Modern fauna (demosponges, bivalves, gastropods, gymnolaemate bryozoans, and malacostracan crustaceans) had its origins in the Paleozoic but its members remained a minor component up to the Permian. During the Triassic, Jurassic, and Cretaceous, members of the Modern fauna originated at a considerably faster rate than members of the Paleozoic fauna, and overall diversity increased steadily throughout the Mesozoic. The End-Cretaceous extinction appears at the family level only as a minor perturbation, and the diversity increase continues throughout the Cenozoic. By the end of the Neogene, family diversity had reached a level almost twice that of the Paleozoic.

The taxic diversity trends observed by Sepkoski were accompanied by several notable trends which relate to the structuring of the marine communities. During the Phanerozoic, the species numbers within open marine assemblages increased through time (denser “species packing”; Bambach 1977), and between the three evolutionary faunas, a significant increase in the number of feeding types and occupied substrate niches (“guilds”) is documented (Bambach 1983). In addition, the proportion of mobile taxa and of predators increased (Bambach et al. 2002). This is paralleled by changes in the spatial structure of marine benthic communities. Animals of Cambrian communities neither extended high above nor burrowed deep into the substrate, but during the mid-Paleozoic to mid-Mesozoic, many suspension feeders evolved that stood highly erect above the substrate and several tiers of epifaunal benthos developed (Ausich and Bottjer 1982; Bottjer and Ausich 1986; Signor 1990). A complex tiering as documented by the trace fossils also developed among the burrowing endobenthos although very deep burrows only became common in the latest Paleozoic.

Following the broad interest in the Phanerozoic diversity trajectories documented for marine animals, various researchers also published figures for other groups of organism (for earlier work see Valentine 1985). For insects, a synoptic curve for the number of families was published that shows a logistic increase in diversity from the Devonian to the recent. The only notable event that led to a significant decrease in the number of families seems to be the end-Permian mass extinction (Labandeira and Sepkoski 1993; Labandeira and Eble in press). Extensive data have also been assembled for tetrapods (Benton 1993). Here, the most significant outcomes are the only slight overall diversity increase from the Devonian to the Late Cretaceous and the huge increase in the number of families after the End-Cretaceous extinction (Benton 1999; Figure 16.3). Plants differ
from the various animal groups in that the family richness over the Phanerozoic follows a much smoother path, and extinctions had little effect on familial diversity (Niklas 1997).

### 16.3.2 Methodological problems

The documentation of Phanerozoic diversity patterns goes back to Phillips (1860) who used data from marine rocks in Great Britain to distinguish between Paleozoic, Mesozoic, and Cenozoic Life and showed that diversity increased throughout the Phanerozoic (see Miller 2000). Renewed interest in these large-scale patterns arose in the 1960s and 1970s, stimulated especially by the pioneering publications of Gregory (1955), Simpson (1960), and Newell (1967). The first diversity patterns published thereafter (Valentine 1969; Raup 1972, 1976a, b; Bambach 1977) showed mainly that this procedure was a very difficult task. The results were, in part, grossly dissimilar, especially with respect to the diversity increase in the Cenozoic. Although a “consensus paper” was published shortly thereafter (Sepkoski et al. 1981), discussions of methodological problems continued till date. The interpretations of Phanerozoic diversity patterns have thus fluctuated between two extremes: it might be the real picture or an artifact of databases affected by different and heavy biases.
There are problems with the taxonomic databases used which can add noise and possibly induce false signals (Signor 1985). Yet two databases at the family level compiled by Sepkoski (1982, 1992) and a team of specialists (Benton 1993) produced very similar and probably reliable results. The “pull of the recent” refers to the fact that the modern biota is much more completely sampled than fossil strata and, therefore, tends to extend the stratigraphic ranges of extant families or species that are known from some time in the Cenozoic through intervals where fossils of those taxa are missing. Yet this bias seems to be a negligible quantity (Jablonski et al. 2003).

Uneven sampling in different regions and in different periods is a fact that cannot easily be corrected for (Smith 2003). Yet it seems that the existing compendia accurately depict the large-scale global diversity trajectories (Adrain and Westrop 2000). Uneven time spans of the sampled intervals (usually stages) are a further problem. To correct for this, longer stages are subdivided, and very short ones are amalgamated (Sepkoski 1996). Taphonomic biases (differential preservation) are usually thought of as occurring randomly through the Phanerozoic, but potential biases can occur, e.g., through taxa that are only rarely preserved (e.g., in extraordinary deposits), producing a local peak in diversity for that stratigraphical interval (Sepkoski 1996). Culling of the data by removal of single occurrences (taxa that are confined to a single time interval) is the common procedure to remove this bias (Sepkoski 1996). Perhaps the most distracting factor when analyzing diversity patterns is that the sedimentary rocks available for study are not evenly dispersed through the Phanerozoic (Raup 1972, 1976b; Peters and Foote 2001; Smith 2001). Yet the diversity trajectories seem to be real and independent of the outcrop area/sediment volume (Miller 2000).

16.3.3 Expansion and equilibrium models

Although the Phanerozoic diversity trajectories and the databases used to construct them have proven to be robust, their interpretations have remained problematic. For the observed increase in familial and generic diversity, three basic models derived from population ecology can be invoked to explain the pattern (Miller 1998; Benton 2001). First, if rates of species origination and extinction were unconstrained by existing diversity and origination exceeded extinction rates, an exponential increase in taxic diversity will result, only interrupted by major extinction intervals. The second, linear model of diversity increase requires the addition of a fixed number of taxa in each unit time. This would require a constant decrease in the rate of evolution (speciation) or a regularly increasing rate of extinction, and this model is generally rejected as...
improbable (Benton 2001). The third, density dependent model assumes that after an initial slow diversity increase a rapid rise occurs, followed by a slowing rate of increase, and finally a plateau (logistic growth).

In a “decoupled logistic” simulation of the three Phanerozoic evolutionary faunas at the family level, which allowed for major mass extinctions, Sepkoski (1981, 1984) achieved a rather good fit to the observed picture. In a variation of the theme, it was proposed that the diversifications from Ordovician to recent were best matched by a series of four simple logistic curves without interaction between evolutionary faunas and only reset by mass extinctions (Courtillot and Gaudemer 1996). Exponential growth (Benton 1995) is currently no more supported for marine animals, but the exponential or nearly exponential increase in several terrestrial groups indicates that these taxa are still far away from reaching a saturation (Benton 2001).

16.3.4 Newer developments

Despite the consensus regarding the Phanerozoic diversity trajectories and the broad agreements between the different databases (Sepkoski 1982, 1992, 2002; Benton 1993), the existing global compilations have some inherent limitations. While major mass extinctions clearly affected diversity on a global scale, the trends in the intervals between these disruptions may show large differences between regions (Miller 1997, 1998, 2001). A new Paleobiology Database is currently under construction (Alroy et al. 2001; Adrain and Westrop 2003; Miller 2003) and assembles locality-specific inventories in the Phanerozoic record. This new approach will allow for a much better correction of taphonomic and sampling bias (Alroy et al. 2001; Bush et al. 2004), will yield new insights into the respective contributions of alpha, beta, and gamma diversity to the Phanerozoic taxonomic richness (Bush and Bambach 2004), and will also ultimately provide us with environment-specific information. It is likely that this will lead to a much more profound understanding of the Phanerozoic diversity patterns.

16.4 Extinctions and biotic recovery: generalities

From the conservation biologist’s point of view, the extinction of even a single species is a catastrophe. From a paleontologists view, species extinction is the normal case and more than 99% of all the species that ever inhabited earth are now extinct (Raup 1991b). Yet there is even a further, more important dimension to this: during the history of life, mass extinctions provided also
huge opportunities for taxa that hitherto played only minor roles, by removing or marginalizing incumbents (Jablonski 2001).

The same databases used in diversity studies can be used to construct plots of extinction (and also origination) intensities over time. Two major points emerge from such an analysis: peaks of high extinction intensities are separated from each other by times of lesser extinction, and overall extinction intensities decline over the phanerozoic (Sepkoski 1996; MacLeod 2003; Figure 16.4). According to Raup and Sepkoski (1982), there were 5 major and at least 18 lesser mass extinctions in the Phanerozoic, and the “big five” major extinctions (Table 16.1) are also recognized in newer studies (Hallam and Wignall 1997; MacLeod 2003; Bambach et al. 2004; Taylor 2004). A “mass extinction” is an event that was (1) nearly global, (2) removed a significant proportion of the existing species (perhaps more than 30%), (3) affected species from a broad range of ecologies, and (4) happened within a (geologically spoken) short time.

For the decline in background extinction intensity, widely disparate explanations were proposed. The change might reflect the general decrease in “volatility” between the three evolutionary faunas (Sepkoski 1981) or secular changes in the geochemistry and nutrition levels in the seas (Martin 1996). Yet it might also just be a taxonomic artifact: through the Phanerozoic, there is a trend toward more species per family/genus. It, therefore, needs more species to become extinct so that the entire family/genus becomes extinct (Taylor 2004).

An extinction event can be either abrupt, or stepped, or gradual. An abrupt or “pulse”-extinction evidently leaves a species no time to adapt or migrate, whereas this would be possible during a gradual or “press” extinction (Erwin 1996b, 2001b). Yet it has proven difficult to establish the exact disappearance of taxa, especially the rare ones. As a consequence of the imperfect fossil record, the observed last appearance of fossil taxa is always “smeared back” in time through a time interval before their actual extinction (Hallam and Wignall 1997; Taylor 2004). This “Signor–Lipps effect” (Signor and Lipps 1982) will lead to the perception of a gradual extinction pattern even if it was abrupt. The “zombie” lineage, that is the unsampled portion of a taxon’s range occurring after the final appearance of the taxon in the fossil record prior to its actual extinction (Lane et al. 2005), can be inferred at some level of probability with statistical methods (Marshall 1990).

Not even the largest mass extinctions acted in a completely random manner. Extinction selectivity can be geographical (e.g., tropical versus nontropical, terrestrial versus marine), taxonomic (different extinction rates among higher taxa, e.g., dinosaurs versus mammals, plants versus animals), or linked to trait (e.g., body size, trophic level; McKinney 1997, 2001). The selectivity patterns seen during a major extinction interval can be the same as those acting during
Figure 16.4
Extinction intensities in the Phanerozoic. Mass extinctions clearly stand out against background intensities, which decreased during the Phanerozoic. After Mac Leod 2003
pre- and postextinction times (“fair game” selectivity) or they can differ (“wanton” selectivity; Raup 1991b). Random survival with respect to trait and taxonomy was termed “field of bullet” selectivity (Raup 1991b).

Major discussions centered on the question whether mass extinctions qualify as a separate class of evolutionary mechanism and are, therefore, different from background extinctions (Jablonski 2001; Bambach et al. 2004). There is certainly a continuity of magnitude, with the “big five” and some Early Paleozoic events occupying the ranks of highest intensities (MacLeod 2003; Wang 2003) although the simple relationship between extinction intensity and mean waiting time (kill curve; Raup 1991a) is no longer accepted because continuity of cause is lacking (Wang 2003). The processes operating during mass extinctions are not the same as those acting during background extinctions (Wignall 2004). Continuity of effects is also no longer tenable as for example the end-Ordovician and end-Devonian mass extinctions both had about the same intensities, but the former had only minor consequences for the structuring of marine communities, while the latter had a profound impact on the ecosystems (Droser et al. 2000). Mass extinctions, therefore, did have a major impact on the evolution of life although most species suffered extinction during times of background extinction intensities (Taylor 2004).

Did the major mass extinctions during the Phanerozoic have a common cause? This subject was treated by several authors, and the mechanisms include impacts (as assumed, e.g., by Raup and Sepkoski 1984), sea-level changes and the spread of anoxia (Hallam and Wignall 1997), large-scale volcanisms (Courtillot 1999), and perhaps global cooling (Stanley 1988). It is not always easy to distinguish between the different mechanisms (MacLeod 2003; Table 16.2). Yet there is now overwhelming evidence that the major mass extinctions had their individual signature, and at the present state of knowledge the search for a common cause does no longer make sense.

It has been advanced already several times that the history of life follows some large-scale cyclic pattern (Fischer and Arthur 1977; Fischer 1984). Yet the notion

### Table 16.1
**Observed (families, genera) and calculated (species) extinction intensities at the five major Phanerozoic mass extinctions**

<table>
<thead>
<tr>
<th>Mass extinction</th>
<th>Families extinct (%)</th>
<th>Genera extinct (%)</th>
<th>Species extinct (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>End-Ordovician</td>
<td>26</td>
<td>60</td>
<td>85</td>
</tr>
<tr>
<td>Late Devonian</td>
<td>22</td>
<td>57</td>
<td>81</td>
</tr>
<tr>
<td>End-Permian</td>
<td>51</td>
<td>82</td>
<td>95</td>
</tr>
<tr>
<td>End-Triassic</td>
<td>22</td>
<td>53</td>
<td>80</td>
</tr>
<tr>
<td>End-Cretaceous</td>
<td>16</td>
<td>47</td>
<td>73</td>
</tr>
</tbody>
</table>

Simplified after Hallam and Wignall 1997.
that Phanerozoic extinction patterns show a periodicity with an interval length of 26 Myr (Raup and Sepkoski 1984) has spurred intensive debates about extinction mechanisms but has received, with increasing accuracy of the geological time scale, little support from other researchers. Currently, few paleontologists would subscribe to the idea that the mass extinctions show any periodicity, despite continued search for possible (mainly extraterrestrial) mechanisms (e.g., invisible companion of the sun, “nemesis,” inducing a comet shower when passing through Oort Cloud; an eccentrical “planet X” doing the same job; the passing of our solar system through the spiral arms of our galaxy, again perturbing bolides in the Oort cloud; movement of the solar system through the galactic plane; Sepkoski 1990).

It comes as a major surprise that a new study indicates again that the diversity of life on earth followed some cyclic pattern, this time with a 62 Ma periodicity (Rohde and Muller 2005). Yet this pattern has yet to be scrutinized, and it might simply be the inevitable outcome of the applied methods.

The phase after a mass extinction usually shows a peculiar fauna/flora that change gradually during the time following the extinction. Some terminology has proven to be helpful for the characterization of these intervals (Kauffman and Erwin 1995; Figure 16.5). The extinction phase is the time where most of the affected taxa had their last appearance. This is followed by the survival phase where some groups not severely affected by the extinction (holdover taxa) already began to diversify. During the survival phase, there is usually a small number of opportunistic taxa that could temporarily spread at the expense of the other fauna.

<table>
<thead>
<tr>
<th>Environmental effect</th>
<th>Comet/asteroid impact</th>
<th>Flood-basalt volcanism</th>
<th>Sea-level change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increased atmospheric particulates</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Increased cloud cover</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Increased greenhouse gases</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Reduced greenhouse gases</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Acid rain</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Global wildfires</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shock heating</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Habitat fragmentation</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Intensification of climate gradients</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enrichment of trace elements</td>
<td>x</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

After MacLeod 1998.
Lazarus taxa (Jablonski 1986) are those that disappear, probably by becoming very rare (and not retreating to refugia; Wignall and Hallam 1999) during the survival phase only to later reappear during the recovery phase when new ecosystems are being established, while Elvis taxa (Erwin and Droser 1993) were those that entered newly the fossil record but mimicked the shape of extinct taxa. The new communities were largely shaped by progenitor taxa, the most prolific during the recovery phase (Kauffman and Erwin 1995; Hallam and Wignall 1997).

It was long recognized that reefs showed, after mass extinctions, a prolonged recovery interval. This “reef gap” (Hallam and Wignall 1997; Stanley 2001; Jablonski 2003), which is evident after all of the “big five” and also after the extinction event after the end of the Lower Cambrian (extinction of the Archeocytha; Copper 1988), probably reflects the longer time it needs to reassemble complex ecosystems (Jablonski 2003).

16.5 Selected case studies

16.5.1 The Cambrian radiation and its prelude

The tracking of the appearance and early diversification of the animals in the fossil record (Figure 16.6) is at present one of the fields in paleontology where extremely rapid progress is made. New input came especially from molecular...
Figure 16.6
Major events during the latest Neoproterozoic and the Cambrian. After Erwin 2001
systematics (Adoutte et al. 2000; Giribet 2002; Pennisi 2003; Halanych 2004; Peterson et al. 2004) and through precise, absolute dating of fossiliferous Late Neoproterozoic and Early Cambrian sections, which allowed the establishment of a reliable chronostratigraphic framework (Martin et al. 2000; Knoll et al. 2004).

Possibly the oldest metazoans are ring- and disk-shaped impressions from the Mackenzie mountains in northwestern Canada, dated as 610–600 Ma and hence of the same age as the terminal Marinoan–Varanger glaciation. This soft-bodied fauna was called Ediacara assemblage I and represents perhaps cnidarian grade animals (Erwin 2001a). Considerably more diverse but younger Ediacaran assemblages are known from a number of localities around the world (Bottjer 2002; Brasier and Antcliffe 2004). The fauna of these Ediacaran assemblages II and III consists exclusively of soft-bodied forms. The body wall of these organisms must have been quite rigid as is evident from the taphonomic behavior of the Ediacaran biota. They are preserved, mostly intact, as impressions at the base of event beds (storm layers, turbidites) that quickly buried them (Erwin 2001a; Bottjer 2002).

Most of the forms are flattened, segmented or quilted, and some of them attain considerable sizes (Narbonne and Gehling 2003). Classification of these organisms has been highly controversial. Traditionally, the various species were assigned to extant groups like medusae, sea pens, polychaete worms, and stem group echinoderms (Wade 1972a, b; Glaessner 1983, 1984; Jenkins 1992), with most of them belonging to cnidarians. Today most researchers believe that the majority of the enigmatic Ediacaran organisms represent “lost constructions” (Seilacher 1989, 1992; Buss and Seilacher 1994; Narbonne 2004, 2005). With the exception of the 555-Ma-old Kimberella, which is now considered to be an ancestral mollusk (Fedonkin and Waggoner 1997; but see Dzik 2003; Budd and Jensen 2004) and therefore the oldest body fossil of a bilaterian, Ediacaran organisms seem to lack a mouth and a digestive tract. Most were immobile, many of them recliners but some also mud-stickers (Seilacher 1999). Nutrient uptake was either through the body surface, or these organisms lived in symbiosis with photosynthetic/chemosynthetic protists or bacteria (McMenamin and McMenamin 1990; Narbonne 2005). Although sponges and cnidarians were most probably also present, predators were notably absent. The immobile organisms thus lived in a “Garden of Ediacara” (McMenamin and McMenamin 1990). The sediment surface was frequently sealed by microbial mats that were not disrupted by larger burrowing animals. Traces including some radular scratch marks are small and rare and are proof of the presence of bilaterians. Claims for evidence of bilaterians earlier than the Ediacaran assemblages (e.g., traces from 1200-Ma rocks in India; Seilacher et al. 1998) are published with some frequency but these are not accepted (Erwin and Davidson 2002).
An exciting new window into Late-Proterozoic life recently opened with the discovery of 600–580-Ma-old fossiliferous phosphorites at Doushantuo, southern China (Xiao et al. 1998). In addition to various algae, microscopic metazoan embryos preserved at the cellular level were found. The affinities of these early cleavage stages are unclear but they might possibly be of bilaterian origin (Knoll and Carroll 1999; Valentine et al. 1999). Newer claims for small bilaterian animals (Vernanimalcula) with preserved coelom and digestive tract from the same locality (Chen et al. 2004) are not widely accepted (Stokstad 2004). The Ediacaran biota largely went extinct at the Precambrian–Cambrian boundary (Brasier and Antcliffe 2004) with perhaps a few survivors into the Middle Cambrian (Conway-Morris 1993).

The base of the Cambrian, which is now dated at 542 Ma (Gradstein and Ogg 2004; Ogg 2004), marks the beginning of an interval which is the most remarkable in the history of life. During the Early Cambrian, a wide range of skeleton-bearing animals made their debut, and within only 20 Myr, almost all phyla with preservable hard parts entered the fossil record (sponges already earlier, bryozoans in the Ordovician, but might have exclusively soft-bodied representatives in Cambrian; Valentine 2002, 2004). This apparent explosion of animal bauplans, which had already puzzled Darwin, is still one of the greatest enigmas, although it has become clear in the last few years that there was a prelude of perhaps tens of million years (Lieberman 2003) seen both among traces and in the shelly fossils, but no lengthy time gap between the Ediacaran assemblages and the lowermost Cambrian (Knoll and Carroll 1999; Erwin 2001; Valentine 2002).

At first we see an increasing size and complexity of burrows/traces across the Precambrian–Cambrian boundary and then an ever increasing diversity during the Early Cambrian (Crimes 1992, 2001; Droser and Li 2001). In parallel, there was a huge diversification in the so-called “small shelly fossils” (Rozanov and Zhuravlev 1992; Conway-Morris 1998, 2001; Grotzinger et al. 2000). The appearance of new animal phyla and other higher taxa was concentrated in an interval between the latest Manykayan and the Atdabanian spanning perhaps 10 Myr, and this interval is called the “Cambrian explosion” and saw even the appearance of phyla that shortly thereafter became extinct (Dzik 1993; Valentine et al. 1999; Conway Morris 2000). By the end of the Cambrian explosion, the phylum-level diversity among marine animals had surpassed that seen in today’s oceans! During this interval, biomineralization was acquired by a multitude of organisms, including mollusks, brachiopods, echinoderms, and chordates (see Lowenstam and Weiner 1989; Bengtson 1994; Zhuravlev and Riding 2001 for an overview). Complex arthropods, like trilobites, enter the fossil record during the Atdabanian stage, and by the Middle Cambrian, a plateau in diversity was reached.
Our picture of the Cambrian explosion is facilitated by several wonderful, exceptional fossiliferous settings (Fossilagerstaetten), including the Early Cambrian localities of Chengjiang (China; Xian-guang et al. 2004) and Sirius Passet (Greenland; Conway Morris 1998), and the most famous Middle Cambrian Burgess Shale (Canada; Briggs et al. 1994). Several very important questions surround debates about the Cambrian explosion. (1) Is it real or just an artifact of the imperfect fossil record? (2) Does the origin of metazoan phyla substantially predate their appearance in the fossil record? (3) If the event was real, what was the triggering mechanism?

The current consensus is that the Cambrian explosion was indeed real and happened in just a 10-Ma interval between the late Manykayan and the Atdabanian (Valentine et al. 1999; Conway Morris 2000; Erwin 2001; Valentine 2002, 2004). Yet the Cambrian explosion does not record the initial split of the metazoans but rather the diversification of the three fundamental metazoan groups into the crown groups and the modern phyla (Douzery et al. 2004; Peterson et al. 2004; Valentine 2004). Ecologically, the Cambrian revolution led to a massive restructuring of the marine benthos. An increasing percentage of animals were burrowers that fed on nutrients within the sediment, or that constructed tubes. As a consequence of the increasing bioturbation (Droser and Li 2001), the sediment became destabilized and the superficial microbial mats were lost. This “agronomic revolution” (Seilacher 1999) led to an uppermost mixed layer, and besides the Ediacara organisms, many immobile recliners and mudstickers like the helicoplacoid echinoderms went extinct (Bottjer et al. 2000; Dornbos and Bottjer 2000; Dornbos et al. 2005).

With respect to the triggering mechanism(s) of the Cambrian explosion, our explanations are still very much speculative. The synchronous radiation of many disparate phyla has led strong support to the idea that there must have been an environmental trigger (physical or biological; Valentine 2002, 2004). Among the hypotheses that are advanced are a significant increase in oxygen in the oceans but this hypothesis is at odds with observations that oxygen reached sufficient levels to sustain metazoan life long before the Early Cambrian (Knoll 1996; but see Hedges 2004). The amalgamation of continental plates followed by a transgression was also cited as a possible trigger (Brasier and Lindsay 2001). There were certainly dramatic changes in the chemistry of the oceans during the latest Precambrian as is evident from isotope studies. Increased nutrient levels in the oceans might have facilitated the radiation (Brasier 1992). Furthermore, a major increase in Ca\(^{2+}\) levels during the earliest Cambrian is documented which certainly facilitated biomineralization (Brennan et al. 2004). One of the most commonly involved biological explanations of the radiation is that skeletons developed in a number of groups as an answer to increasing predator pressure.
(Conway Morris 2001; see also Stanley 1973) and perhaps also the evolution of animal vision (Parker 2003), thereby allowing new adaptational opportunities. The Cambrian saw indeed the appearance of many carnivores and scavengers, and this trophic group accounted for up to 25% of the species in the Middle Cambrian (Burzin et al. 2001). Furthermore, the development of planktonic habits in a range of groups greatly increased the complexity of food webs and was also responsible for animal diversification (Butterfield 2001).

Many paleontologists were deeply impressed by the wide variety of different body plans, especially among arthropods, that appeared during the Cambrian explosion. The number of taxa known from the famous Cambrian Lagerstätten that do not fit easily into any classification scheme based on living animals is extraordinary. It was claimed that morphological disparity among Cambrian animals was even higher than that seen today (Gould 1989, 1991) and some extraordinary evolutionary mechanisms acted during the Cambrian explosion. Measuring morphological disparity is a relatively new field and seems to be a promising avenue to complement measures of taxonomic diversity (Foote 1997; Wills 2001). Detailed analyses of arthropods showed that disparity among the Cambrian forms was not higher than among modern species and the level of appendage specialisation is much higher today than in the Cambrian (Briggs et al. 1992; Wills et al. 1994). Special evolutionary processes do, therefore, not seem to have operated during the Cambrian explosion. Yet it remains nevertheless remarkable how early and synchronous the multitude of higher animal taxa appeared in the fossil record.

16.5.2 The Ordovician radiation

During the Ordovician, taxic diversity at the family and generic level reached a new maximum which was three to four times that of the Late Cambrian (Miller 1997, 2001). While during the Cambrian explosion numerous phyla and classes representing basic body plans originated, the Ordovician radiation was manifested by an unprecedented burst of diversification at lower taxonomic levels. According to the global picture of Sepkoski (1979, 1997), the Cambrian fauna (e.g., the trilobites) declined while the articulated brachiopods, the crinoids, stenolemate bryozoans, and other members of the Paleozoic fauna showed a sharp increase. To a lesser degree, members of the modern fauna like gastropods and bivalves also diversified (Miller 2001).

The almost exponential increase in diversity was much more rapid during the Ordovician than at any other time of the Phanerozoic (Sepkoski 1997; Webby 2004). This global diversity increase seems to have been the combined result of an
increase in $\alpha$- (within community), $\beta$- (between communities), and $\gamma$-diversity (biogeographic differentiation between faunal provinces; Webby 2004). New ecological guilds appeared, and the spatial organization of the benthic communities became considerably more complex (Bottjer and Ausich 1986; Droser and Bottjer 1989, 1993). The Ordovician was also a time of marked shift in the reef biota. Whereas Early Ordovician reefs were mainly built of microbial mats and stromatolites, large metazoan-dominated framework reefs had developed and spread on all continents by the end of the Ordovician (Webby 2004). The diversification and spread of metazoan reefs was accompanied by an increase in bioeroding organisms colonizing reefs and hardgrounds ("bioerosion revolution"; Wilson and Palmer 2001).

Yet when the global picture is dissected at finer taxonomic, geographic, and environmental levels, some surprising results emerge (Miller 1997). The diversification seemed not to follow a global trajectory but to respond much more to local conditions in paleogeography, sedimentary environment, and orogenetic activity. The general decline of Sepkoski’s Cambrian fauna is an oversimplification. The “smooth” global diversity increase during the Ordovician is, thus, the aggregate record of many regional patterns which were for the most part abrupt (Miller 1997, 2001, 2004; Westrop and Adrain 1998).

Not a single reason but rather a combination of paleogeographic, sedimentologic, geochemical, and perhaps intrinsic factors seems to be responsible for the diversity increase during the Ordovician. The paleogeographic situation was one of a highly fractured continental crust with many small continents. Together with a very high sea-level stand and increasing nutrient inputs, this provided extensive shelf areas in many different parts of the world and, therefore, a multitude of colonizable areas (Miller 1997). There is also a good correlation between diversity increase and orogenic activity. Throughout the Ordovician and the remainder of the Paleozoic, there was clearly a trend toward increased primary productivity and away from oligotrophic conditions (Martin 1996).

16.5.3 The Late Ordovician mass extinction

With 26% of the marine families, up to 60% of the genera and an estimated 85% of all the species becoming extinct, the Late-Ordovician mass extinction was one of the most severe extinction episodes in the Phanerozoic, surpassed in its magnitude only by the end-Permian mass extinction. Its extent was global, and it affected nearly all benthic and pelagic groups (Hallam and Wignall 1997; Brenchley 2001; Sheehan 2001b; Figure 16.7).
Figure 16.7 Extinction patterns during the Late Ordovician mass extinction. Modified after Hallam & Wignall 1997.
During the Late Ordovician, greenhouse conditions prevailed but during the latest stage (Hirnantian), a short glaciation of approximately 0.5-Ma duration occurred (Brenchley 2001; Sheehan and Harris 2004). The glaciation probably started when Late Ordovician orogenic activities led to extensive exposure and weathering of silicate terrains and, therefore, CO$_2$ consumption. Under falling CO$_2$ levels, ice sheets started to grow and albedo feedback led to an extensive Gondwana glaciation. The ice cover in turn inhibited silicate weathering, and CO$_2$ levels rose again. After a threshold was reached, greenhouse conditions returned and the ice caps melted quickly (Kump et al. 1999). It is now well documented that the mass extinction during the Late Ordovician was a two-phase event and the first extinction pulse at the base of the Hirnantian *extraordinarius* zone corresponded to the initiation of the Gondwana glaciation and the second pulse in the first third of the *persculptus* zone to the rapid decay of the polar ice cap.

During the first extinction event, various groups suffered heavy losses. Thereafter, an impoverished and remarkably cosmopolitan benthic fauna was present on most shelves from high latitudes to the tropics. This so-called *Hirnantia* fauna can be regarded as a cool-adapted, opportunistic community that spread after the extinction removed the hitherto dominant species (Brenchley 2001; Sheehan 2001b; Jia-Yu et al. 2002). Primary productivity was obviously much reduced during the Hirnantian.

The second extinction phase was again sharp and hit most of the groups that suffered already during the first extinction pulse. It eliminated much of the benthic *Hirnantia* fauna (Hallam and Wignall 1997), but the recovery interval did not last very long. It is most surprising that the faunal turnover during the Late Ordovician mass extinction was accompanied by very little ecological changes, and the structure of Silurian communities is remarkably similar to those of the Late Ordovician (Droser et al. 2000; Bottjer et al. 2001; Sheehan 2001b).

The causes of the two extinction pulses were certainly linked to the rapid onset and the later abrupt termination of the Gondwana glaciation during an otherwise warm climatic mode (Fortey and Cocks 2005). Global cooling of the oceans of perhaps as much as 8°C together with a loss of benthic habitat due to regression might have, in part, been responsible for the first extinction (Berry and Boucot 1973; Brenchley et al. 1994; Armstrong 1996). Yet pelagic forms also suffered heavy losses, and changing circulation patterns in the oceans were probably crucial (Hallam and Wignall 1997). The widespread deep anoxic waters, the extensive dysoxic zone, and the nutrient-rich surface waters vanished during the onset of the glaciation when cold, deep water led to intensified ocean circulation. During the second pulse, the termination of cold, deep water production led again to widespread stratified oceans with
anoxic deep and intermediate dysoxic waters. Black shales accumulated again, and the transgressing dysoxic waters eliminated most of the benthic *Hirnantia* fauna and other benthic organisms not resistant to oxygen-poor conditions (Hallam and Wignall 1997).

### 16.5.4 The Late Devonian mass extinction

No less than 11 global extinction events were identified in the Givetian through Famennian stages (Walliser 1996) but only the lower and upper Kellwasser events (Frasnian–Famennian boundary) and the Hangenberg event (uppermost Famennian) amount to a magnitude that deserve the name mass extinction (Hallam and Wignall 1997). Yet the stress imposed by the many smaller events, especially the earlier Thaganic event in the uppermost Givetian, was probably crucial for the overall extinction patterns in maintaining a high level of environmental stress throughout the Late Devonian.

At least 70% and perhaps as many as 82% of the marine species became extinct during this time period (McGhee 1996, 2001; [Figure 16.8](#)). Among the groups that were most severely hit were the reef builders (Fagerstrom 1994; Hallam and Wignall 1997; Copper and Scotese 2003), but other benthic organisms, especially tropical families, also suffered heavy losses during both major crises. The toll was no less severe in planktonic and nektonic groups, and the armored agnathans and the placoderms went completely extinct (Hallam and Wignall 1997).

Quite important for the elucidation of the possible causes of the Late Devonian mass extinctions is that their selectivity differed between the different events. The Taghanic event affected mainly benthic taxa from low-latitude, shallow-water environments (Hallam and Wignall 1997). The Kellwasser events also affected mainly warm-water species as well as planktonic and pelagic groups. During the Hangenberg crisis, it was the planktonic and nektonic groups that were most severely hit while the benthic groups showed a better survival than at the Frasnian–Famennian boundary.

The extinction patterns in the Late Devonian were highly complex and a result of several mechanisms spread over a time period of more than 10 Ma, with the most severe perturbations concentrated at the Frasnian–Famennian boundary and in the latest Famennian (Sandberg et al. 2002). Global cooling of the oceans (Copper 1986) was certainly one of the main causes of the extinctions (McGhee 1996, 2001), although the Gondwana glaciation only started in the Late Famennian (Caputo 1985; Algeo and Scheckler 1998). Yet beside the cooling, frequent sea-level changes including both eustatic rises associated with spreading anoxic waters and
Figure 16.8
Extinction patterns during the Late Devonian mass extinction. Modified after Hallam & Wignall 1997
regressions responsible for habitat loss are also well documented (Sandberg et al. 2002). The Late Devonian was also a time of increased impact frequency. Well-dated craters, shocked minerals, and microtectites as well as iridium anomalies are known from different continents (McGhee 1996, 2001; Sandberg et al. 2002), and these impacts certainly increased the environmental stress.

### 16.5.5 Devonian to Carboniferous expansion of land flora

The earliest vascular land plants are known from Late Silurian strata. Colonization of the terrestrial environment, at least in moist lowlands, obviously happened rather quickly, and within only 45 Ma, all major land-plant lineages and organizational grades (except for flowering plants) developed (Niklas 1997, 2004; Willis and McElwain 2002). All the morphological adaptations including a protective outer covering (waxy cuticula) against desiccation, stomata to allow gas diffusion, specialized tissues for the transport of liquids, and rigid cell walls (Chaloner 2003) developed in the Devonian, although accompanied only by a modest diversification at the species level (Niklas 1997; Willis and McElwain 2002). By the end of the Devonian, the terrestrial environment saw the first globally distributed forests with large trees (Chaloner 2003). This resulted in a huge increase in biomass that culminated in the Late Carboniferous. Much of this organic material was not decomposed and recycled but instead buried in moist anoxic soils (acidic swamps; Chaloner 2003).

This expansion of the land flora had a profound impact on earth’s environmental conditions. Most important was the removal of large quantities of CO₂ from the atmosphere through photosynthetic carbon fixation. The most widely accepted models for Phanerozoic CO₂ show a sharp decrease from about 15 times the present level (15 PAL) at the beginning of the Devonian to 10 PAL at the Devonian–Carboniferous boundary and a further decrease to less than 2 PAL in the Late Carboniferous (Berner 1998; Royer et al. 2000). With the massive drop in available CO₂, a high stomatal density became crucial, and the laminate leaf rapidly became widespread (Beerling et al. 2001) although at the cost of higher water loss through transpiration (Chaloner 2003).

During the Devonian, the increase in the sizes of the trees was accompanied by increasing depth and complexity of the roots (Algeo and Scheckler 1998). This in turn accelerated silicate weathering and led to a further drawdown of carbon as bicarbonate into rivers and ultimately into the seas (Kump et al. 2004). The most obvious consequence of this huge decrease in atmospheric CO₂ was the onset of the Gondwana glaciation. In the Late Devonian, the southern continents were
assembled near the South Pole, and the first polar ice caps developed in the latest Devonian. However, the main phase of the Late-Paleozoic glaciation started in the Early Carboniferous. With its duration of more than 80 Ma (Crowley and North 1991; Frakes et al. 1992), well into the Permian, this glaciation was by far the longest and also the latitudinally most extensive of the Phanerozoic ice ages. Yet it was not associated with any major mass extinction pulse.

The very high productivity of the plants during the Devonian and Carboniferous also led to a significant increase in atmospheric oxygen, with 30–38% O$_2$ in the Late Carboniferous (Berner 1999; Berner et al. 2003; Bergman et al. 2004). This seems to have had yet another impact on life. With increasing oxygen partial pressure, the diffusive flux is increased considerably, allowing the evolution of gigantic sizes (Graham et al. 1995; Dudley 1998), most notably among terrestrial arthropods and amphibians (Briggs 1985; Graham et al. 1995). For the giant flying insects, the higher density of the atmosphere might also have played a role (Dudley 1998, 2000).

16.5.6 The end-Permian mass extinction and subsequent recovery

The mass extinction at the end of the Permian was recognized already for a long time as the most severe of all the Phanerozoic perturbations (Phillips 1860). The radical faunal change associated with this biotic crisis was the reason to distinguish between the Paleozoic below and the Mesozoic above. More than 50% of marine and terrestrial families went extinct and an estimated 80–96% of all the species. Until recently (Erwin 1990, 1993), the end-Permian mass extinction was seen as a protracted crisis, which lasted for approximately 10 Ma. Newer research has shown that there were actually two discrete events (Figure 16.9). The first occurred in the latest Guadalupian (terminal Middle Permian) and affected only some groups in a more gradual way (Hallam and Wignall 1997). The event at the very end of the Permian was apparently of quite short duration (probably less than 0.1 Ma; Erwin et al. 2002). It is this interval that has been called “the mother of all extinctions,” “the great dying,” or the “Paleozoic nemesis” (Erwin 1996; Benton 2003). This massive crisis affected all groups of organisms, both in the seas and on land (Benton and Twitchett 2003).

In the Middle Permian, the seas were teeming with life and many different faunal provinces can be distinguished. Highly diverse stromatoporoid-coral reefs were widely distributed. On soft- and hard-bottoms, rich communities dominated by brachiopods and echinoderms flourished, and in the water-column, numerous ammonoids and various fish groups had attained a high diversity.
Figure 16.9
Extinction patterns during the end-Permian mass extinction. Modified after Hallam & Wignall 1997
On land, insects had reached their highest diversity in the Paleozoic and tetrapod communities were probably as complex as modern mammal communities (Benton 2003). Plants were also highly diverse and distributed in different biogeographical provinces.

During the first extinction event, which occurred at the end of the Middle Permian (Guadalupian), some groups were affected both on land and in the seas but none became entirely extinct (Hallam and Wignall 1997). The reasons for this first extinction pulse are not well understood, but global cooling has been cited as the underlying cause (Stanley 1988; Hallam and Wignall 1997).

The second and by far the more severe extinction pulse near the Permian–Triassic (P–Tr) boundary affected all the taxonomic and ecologic groups, both in the seas and in the terrestrial environment. Among the larger groups that completely went extinct in the seas were the rugose and tabulate corals, fenestrate bryozoans, and orthid brachiopods. On land, glossopterids and cordaitales were suddenly replaced by a low-diversity conifer-lycopod-fern assemblage with little provinciality. Palynological samples from immediately above the P–Tr boundary are dominated by fungal spores which normally account for only a small proportion of the pollen and spores. This “fungal spike” (Eshet et al. 1995; Visscher et al. 1996), similar in its magnitude to the “fern spike” at the Cretaceous–Tertiary boundary, might indicate vast areas of rotting plants which were decomposed by fungi (Hallam and Wignall 1997; Benton and Twitchett 2003). A wide range of tetrapods, among them the hitherto dominant pareiasaurs, went abruptly extinct, and the Early Triassic vertebrate faunas were completely dominated by the single genus *Lystrosaurus* (Hallam and Wignall 1997; Benton and Twitchett 2003; Ward et al. 2005). Both remarkable were the magnitude (up to an estimated 96% of the marine species) and the extraordinary long recovery interval. It took almost 100 Ma for family diversity to reach preextinction levels and almost 10 Ma for complex ecosystems like reefs to become established again (Benton and Twitchett 2003). The first communities that appeared during the recovery interval are composed of a remarkably cosmopolitan, opportunistic fauna of thin-shelled bivalves (e.g., *Claraia*) and lingulid brachiopods (Hallam and Wignall 1997). Burrowing organisms were almost completely absent, and disaster taxa like stromatolites became locally abundant (Schubert and Bottjer 1992). Lazarus taxa were especially common among the gastropods, and most of them were small (“Lilliput”-effect; Twitchett 2006). The long-term consequences for seafloor communities were the replacement of the hitherto dominant epibenthic sessile suspension feeders by a vagile, epi- and endobenthic, mollusk-dominated fauna (Hallam and Wignall 1997). The recovery interval after the end-Permian mass extinction is much longer than for any other extinction event and indicates that the ecosystems were almost completely devastated and
severe environmental perturbations continued through the Lower Triassic (Twitchett 1999; Payne et al. 2004).

The scenario for this mass extinction and its likely causes have received much attention in the last decade. All the evidence indicates that the mass extinction occurred during a phase of marine transgression and severe global warming. The catastrophe probably started with the release of huge amounts of CO2 into the atmosphere, first through volcanic eruptions in South China (Emeishan flood basalt province; Lo et al. 2002) and perhaps also through coal oxidation (Hallam and Wignall 1997), later through vast eruptions in Siberia (Siberian traps; Courtillot 1999; Benton and Twitchett 2003; but see Wignall 2001a). This led to global warming and in the seas to decreased ocean circulation and oxygen depletion (Wignall and Twitchett 1996). With further increasing CO2 levels, methane hydrates began to melt and released large quantities of methane, which first acted as greenhouse gas and later was oxidized to CO2. Through this positive feedback, a “runaway greenhouse” developed, which went out of control after some threshold was reached (Benton and Twitchett 2003). The seas flooding the shelves became anoxic, perhaps even sulfidic (Kump et al. 2005), and killed most of the benthic and pelagic organisms. On land, the vegetation suffered a severe deterioration with equally devastating consequences for the animals, which also experienced hypoxic stress (Huey and Ward 2005).

Inevitably, many additional or alternative explanations have been presented. A major global regression during the terminal Permian was long a popular explanation for the extinction but this is no longer tenable. There was clearly a transgression during the P–Tr boundary interval (Hallam and Wignall 1997). Darkening and global cooling with a collapse in photosynthesis was also proposed as extinction cause (Campbell et al. 1992), but all the evidence points to global warming at the end of the Permian. Older suggestions include brackish oceans and an increase in cosmic radiation. The claim for evidence of an impact is relatively recent (Becker et al. 2001, 2004). However, the impact hypothesis is now generally considered invalid (Erwin 2003). The end-Permian mass extinction thus seems truly “home-made” (Benton and Twitchett 2003).

16.5.7 End-Triassic extinction

The extinction at the end of the Triassic (Figure 16.10) is recognized as one of the “big five” Phanerozoic mass extinctions but documenting the exact timing and the causes of biotic overturn has proven difficult. There was a widespread regression at the end of the Triassic and marine sections which span the Triassic–Jurassic (Tr–J) boundary are known from only a few localities (Hallam and Wignall 1997).
Many marine groups suffered a dramatic and sudden decrease in diversity at the end of the Rhaetian, but others witnessed a major reduction in diversity already during the Norian and earlier Rhaetian (Teichert 1990; Hallam and Wignall 1997; Hallam 2002; Tanner et al. 2004). Terrestrial plant extinction intensities seem to have been concentrated at the Tr–J boundary, and the boundary layer contains an unusually high fern spores/pollen ratio (Olsen et al. 1990, 2002).

Among terrestrial vertebrates, a major extinction is undisputed, but the main turnover occurred near the Carnian–Norian boundary (Benton 1994; Lucas 1994). Instead of an abrupt extinction, there is, thus, a rather complex pattern, and no single catastrophe seems to be responsible for the terminal Triassic turnover. Global climate changes could explain the decline in terrestrial diversity during the
Late Triassic (Tanner et al. 2004). Habitat loss and changing substrates associated with the regression in the latest Rhaetian, followed shortly thereafter by a transgression at the Rhaetian–Hettangian boundary, might have in part been responsible for the observed pattern in the marine realm. Contrary to earlier beliefs (Hallam and Wignall 1997), the transgression at the Tr–J boundary does not seem to have been accompanied by widespread anoxic waters on the shelves (Hallam and Wignall 1999; Tanner et al. 2004), but such an event might have occurred earlier, during the Norian–Rhaetian transition (Tanner et al. 2004). Extensive and widespread volcanism related to the rifting of Pangea around the North Atlantic at the Tr–J boundary was only recently recognized. The outgassing of CO₂ from this Central Atlantic Magmatic Province (CAMP; Marzoli et al. 1999) might have had truly deleterious effects like enhanced seasonal fluctuations and an increase in the number and severity of hot days as well as a decrease in ocean water oxygenation (Huynh and Poulsen 2004). An impact scenario (Olsen et al. 2002) is largely dismissed today because the extinction pattern is not a sudden, catastrophic one, various impact craters have been dated as Carnian–Norian, and claims for significant iridium anomalies and shocked quartz could not be verified (Hallam 2002; Tanner et al. 2004).

16.5.8 Mesozoic marine revolution

The term “Mesozoic marine revolution” (MMR) (Vermeij 1977) refers to the idea that during the Mesozoic a profound reorganization in the marine communities led to a significant increase in predation pressure and prey species developed various adaptations (thicker shells, spines, behavioral responses) to cope with this increasing pressure (“arms race” or “escalation”; Vermeij 1987; Harper 2003, 2006). It is undisputed that during the Mesozoic, especially during the Jurassic and Cretaceous, the number of marine grazers as well as durophagous and drilling predators increased considerably (Vermeij 1977, 1987).

This rise in predatory groups was accompanied by profound changes in marine benthic communities. The epifaunal guilds like stalked crinoids and brachiopods, which were so characteristic of Paleozoic communities, vanished from shallow-shelf environments, and those epifaunal species that did persist in shallow water do show a high frequency of regeneration and, therefore, predator attacks (Vermeij 1987). A marked shift toward infaunal life modes is documented in post-Paleozoic echinoids, gastropods, and especially bivalves (Stanley 1977; Thayer 1983; Vermeij 1987), although this shift predates the appearance of most shell crushers (Harper 2003). The most conspicuous changes during the MMR occurred in shell architecture. Overall, the shells became sturdier, more highly armored, and
developed spines, ribs, and thickened and narrowed apertures (Vermeij 1977, 1987; Ward 1981, 1983). At least among gastropods, shell-repair scars became much more frequent, pointing again to increasing predation pressure.

According to Vermeij (1987), it was the biological interactions (competition, predator–prey relations) that led to the evolution of these long-term trends. Yet the biological evolution toward increasing predator-resistant shells might also have been facilitated by changes in the abiotic conditions. Extensive volcanism, which in turn augmented water temperature, high nutrient levels, and a high sea-level stand all facilitated the production of energetically expensive massive shells (Vermeij 1995) and perhaps as well the general increase in diversity and increasing “fleshyness” of the fauna throughout the Mesozo- and Cenozoic (Bambach 1993). Secular changes in oceanographic and geochemical conditions most certainly spurred other important changes in the marine biota during middle and late Mesozoic times. Planktonic foraminifers appeared in the Middle Jurassic and became numerically important during the Cretaceous. Coccoliths are known since the Triassic, but they became widespread and abundant during the Late Jurassic and the Cretaceous. This rise in planktonic calcifiers was perhaps facilitated by the intensified bioturbation of the seafloor which effectively recycled nutrients (Kelley and Hansen 2001) but had in turn tremendous effects on the carbon cycle and the CaCO₃ saturation of the oceans (Ridgwell 2005).

16.5.9 End-Cretaceous mass extinction

This is certainly the most widely known and probably also the best investigated of the major mass extinctions, simply because the popular (nonavian) dinosaurs disappeared at the Cretaceous–Tertiary (K–T) boundary. Yet it is, with a 16% loss of the families, a 47% loss of the genera, and an estimated loss of at least 70% of the species in the marine realm, the least severe among the five major mass extinctions in earth history (Jablonski 1994; Hallam and Wignall 1997; Figure 16.11). Some marine groups disappeared completely at the end of the Cretaceous (e.g., the large marine reptiles), others suffered heavy losses (especially planktonic groups), but there were also groups that exhibited little or no reduction over the last Cretaceous to the lowermost Paleogene (MacLeod et al. 1997; Norris 2001).

On land, plants suffered evidently little long-term reduction (Niklas et al. 1983; Niklas 1997; Willis and McElwain 2002), but in many sections a short proliferation of ferns at the expense of angiosperms (“fern-spike”) is documented. Among the tetrapods, amphibians, turtles, crocodilians, and eutherian mammals were largely unaffected by the K–T boundary event, whereas lizards and marsupials suffered heavy losses (Archibald and Fastovsky 2004). For the
Figure 16.11
Extinction patterns during the K-T mass extinction. Modified after Hallam & Wignall 1997; MacLeod et al. 1997

<table>
<thead>
<tr>
<th></th>
<th>Late Cretaceous</th>
<th>Paleocene</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cenomanian</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Turonian</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Coniacian</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Santonian</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Campanian</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Maastrichtian</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Danian</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Selandian</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Thanetian</td>
<td></td>
</tr>
</tbody>
</table>

- coccoliths, planktonic foraminifers
- corals
- brachiopods
- rudist bivalves
- ammonites
- teleosteanas
- amphibians
- turtles
- ichthyosaurs
- plesiosaurs
- lepidosaurs
- crocodiles
- pliosaurs
- dinosaurs
- birds
- mammals
ornithischians and the (nonavian) saurischians, this event was of course the end of a long era. Land-dwelling species were more severely hit than freshwater inhabitants, endothermic tetrapods (including ornithischians and saurischians) more than ectothermic, and larger more than smaller species (Archibald 1996; Archibald and Fastovsky 2004).

Evidence for both a gradual decline in dinosaur species richness as well as for a catastrophic and (geologically spoken) instantaneous extinction of the dinosaurs was presented (Hurlbert and Archibald 1995; Archibald and Fastovsky 2004), but the most recent data indicate again that the dinosaurs went abruptly extinct during the time of their highest diversity (Fastovsky et al. 2004). The picture is equally complicated in marine exposures in which again evidence for both a gradual as well as a sudden extinction was presented (see MacLeod et al. 1997). Yet resampling of formerly investigated sections extended taxonomic ranges upward and the reported gradual decline of many groups might well be a consequence of the Signor–Lipps effect (Ward 1990).

Among the possible causes for the K–T mass extinction that are still considered today are volcanism, climatic fluctuations, and marine regression, and an asteroid impact (Benton 1990). There is indeed overwhelming evidence that the earth was hit by a major asteroid perhaps 10 km in diameter that had a devastating impact on earth’s life. This evidence includes molten sediment particles (glass spherules), shocked quartz grains, and a worldwide recognized enrichment in iridium in K–T boundary layers (Alvarez et al. 1980, 1995). The impact hypothesis received further support with the discovery of a 65-Ma old impact crater (Chicxulub) on the Yucatan peninsula, Mexico (Hildebrand et al. 1990). Yet it has also been demonstrated that the latest Cretaceous was a time of major climatic fluctuations (Skelton 2003) and the pronounced marine regression at the end of the Maastrichtian was recognized already a long time ago. In addition, intensive volcanism, which spanned less than 2 Ma over the K–T boundary, is documented from the so-called “Deccan Traps” (Courtillot 1990, 1999). These represent immense outpourings of lava in what is today India and must have had a profound impact on the biosphere (Kelley 2003).

Currently there are two schools of thought to explain the K–T mass extinction. According to the gradualistic, multiple causes–scenario (Archibald 1996; see also Archibald and Fastovsky 2004), the climatic fluctuations and the marine regression near the end of the Cretaceous changed profoundly the available habitats in both the terrestrial and marine environments. This biotic stress led to a gradual decline in the dinosaurs and other terrestrial vertebrates. Likewise, the regression also imposed a major stress on the marine animals by reducing the available shelf environments. Further stress was imposed by the Deccan Trap volcanism that erupted large amounts of dust into the atmosphere, with general
cooling of the globe, drying of many terrestrial ecosystems, and slowing of the photosynthetic activity as the most likely consequences. The asteroid impact at the K–T boundary with its devastating environmental consequences is not disputed in the gradualist camp but is seen merely as the “last strike” that led to the collapse of already weakened ecosystems and extinguished animal groups that were already in decline.

Those researchers that favor a single cause for the mass extinction at the K–T boundary emphasize the almost apocalyptic effects a bolide impact would have (Alvarez et al. 1995). This impact ejected considerable amounts of molten rock particles and dust into the atmosphere. Furthermore it produced huge tsunami-type waves that devastated the coastal plains, and an immense fireball ignited vast wildfires. The dust particles would remain in the atmosphere for months, perhaps even years, leading to global cooling and darkening. Photosynthesis came almost to a halt, at least in plants adapted to higher light intensities. The fern spike recorded from many terrestrial boundary sections testifies to this sudden decline in higher plants and the spread of the ferns, which could cope with darker conditions. As a consequence of the collapse of the ecosystems, first the consumers and then the carnivores died out within years. Those animals that did survive were preferentially small, unspecialized, opportunistic species that could feed on a variety of diets. That ecosystems were also severely and almost instantaneously hit in the marine realm is indicated by the patterns of the stable carbon isotopes across the boundary. These indicate an almost lifeless ocean after the K–T boundary (“Strangelove Ocean”; Hsu and McKenzie 1985; Zachos et al. 1989).

The normal succession in Caribbean coastal sections agrees well with this scenario (after Alvarez et al. 1995): above the Maastrichtian limestone, larger airborne particles (microtectites, glass spherules) were deposited first, then come tsunami deposits containing reworked Maastrichtian limestone and charcoal (from wildfires), then dust-borne iridium and shocked quartz, and finally we see a return to normal sedimentation. Yet in some sections, the succession of the different layers is incomplete, and sometimes there are multiple ejecta and iridium layers (some of them perhaps reworked). According to a new scenario, three impacts within 0.3–0.4 Ma around the K–T boundary best explain the patterns and all contributed to the mass extinction (Keller et al. 2003). The one producing the Chicxulub crater predated the K–T by 0.27 Myr.

16.5.10 Eocene–Oligocene transition

In a diagram of Phanerozoic extinction intensities, the Eocene–Oligocene transition period barely stands out as an important event, and yet this time marked the
most significant episode since the extinction of the dinosaurs (Prothero 1994). The Eocene was a time of warm temperatures, with widespread tropical forests, archaic mammals, and reptiles occurring above the arctic cycle (Prothero 1994; Ivany et al. 2003). Within a time period of 10 Myr, this “greenhouse” world shifted to “icehouse,” with decreased global average temperatures and markedly increasing seasonality, accompanied by major shifts in the biota of terrestrial environments and the seas (Berggren and Prothero 1992).

The record in the marine environment is one of a rather gradual turnover. A displacement of warm-adapted taxa by invading high-latitude forms is observable in both planktonic and benthic taxa and was obviously accelerated around the Middle-Late-Eocene (Bartonian–Priabonian) boundary and within the Early Oligocene (Rupelian) (Berggren and Prothero 1992). As much as 90% of the genera disappeared in some groups (Hallam and Wignall 1997). Fishes survived this period almost unaffected while among the whales, the more basal archaic whales were replaced by the first modern toothed and baleen whales (Prothero 1994).

On land, the flora changed between the Middle Eocene and Early Oligocene from widespread forests to open shrublands (perhaps with the first grasses), and the leaf record as well as the amphibian and reptile fauna show a marked cooling and drying trend (Berggren and Prothero 1992; Prothero 1994). Among the mammals, a fundamental difference between the Eocene and Oligocene faunas, the “Grand Coupure,” has long been recognized (Stehlin 1909; Legendre and Hartenberger 1992). In Europe, there is a gradual replacement of archaic by modern mammals throughout the Late Eocene and Early Oligocene with a pronounced peak of extinctions at the Grande Coupure (Hooker et al. 2004). This latter event is now dated as Early (but not earliest) Oligocene (Prothero 1994). Almost all new taxa are immigrants from Asia, which could reach Europe after the closure of the Turgai strait in the Ural region. Similar trends are observable on the other continents (Prothero 1994).

High-resolution oxygen isotope studies over the past decades have considerably helped to clarify the climatic evolution from the Early Eocene to the Early Oligocene. Global temperatures reached a peak between 52 and 50 Ma, during the so-called Early Eocene Climatic Optimum (EECO). This was followed by a 17-Ma–long trend toward cooler conditions, with a dramatic increase in δ18O at the Eocene–Oligocene boundary (Zachos et al. 2001). This latter shift in the oxygen isotope values reflects not only cooling but also a significant increase of the Antarctic ice cap. The first ephemeral ice sheets appeared already during the Late Eocene and were probably the result of both declining atmospheric CO2 levels (DeConto and Pollard 2003) and an increasing thermal insulation of Antarctica. Plate tectonic movements had led to the separation of Australia
and Antarctica and to the opening of the Drake Passage between Patagonia and Antarctica. By the end of the Eocene, the passage south of Tasmania was deep enough that the circum-Antarctic cold deep current could become established, and the Antarctic ice cap rapidly grew (Prothero 1994; Ivany et al. 2003; Kennett and Exon 2004). Yet on land, the extinctions associated with the Eocene–Oligocene transition were not simply the result of global cooling, increased seasonality, and increasing aridity but also a consequence of large-scale migrations. As an alternative or perhaps additional cause of the terminal Eocene turnover, impacts by comet showers were proposed (Poag et al. 2003). For the Siberian Popigai (100 km) and the Chesapeake Bay crater (85 km) as well as for the ejecta layer at the Massignano type section in Italy, a corresponding age has been established. Yet the biotic turnover patterns do not seem to correspond to an impact scenario (Ivany et al. 2003).

16.5.11 Pleistocene and modern extinctions

During the Pleistocene, between 1.8 Ma and 0.01 Ma, the earth witnessed large climatic fluctuations. Both plants and animals on all continents showed large shifts in their geographical distribution during this period but the extinction levels were not above background values (Alroy 1999). This is true for such disparate taxonomic groups as insects, amphibians, reptiles, birds, and mammals. Yet there is one major exception to this rule: the so-called mammalian megafauna (animals > 44 kg). Fifty thousand years ago, more than 150 genera of this megafauna were populating the continents, but 10,000 years ago, at least 97 of these genera were extinct (Barnosky et al. 2004). There have been continued debates over whether these extinctions were caused mainly by environmental changes associated with climatic fluctuations, or were the consequences of human impacts. It is undisputed that humans were responsible for the extinction of large mammals and large birds on islands such as Madagascar, Antillean, Mediterranean, East Asian Islands, and New Zealand (Barnosky et al. 2004; Burney and Flannery 2005). Here hunting and habitat fragmentation led to extinctions even in the absence of climatic change.

An important aspect is that mammalian megafaunal species on all continents became extinct but both magnitude and timing of the extinctions differed between continents (Roy 2001; Barnosky et al. 2004). The extinctions were most severe in Australia where 14 out of 16 (88%) giant marsupials succumbed. In addition, all seven genera of megafaunal reptiles and birds went completely
extinct (Barnosky et al. 2004). Humans arrived on that continent somewhere between 71,500 and 44,200 years ago, and most megafaunal species became extinct before 40,000 years ago (Barnosky et al. 2004). In North America, 33 genera (72%) went extinct (Roy 2001) within a short time interval between 11,500 and 10,500 years ago, closely correlating with the arrival of Clovis-style hunters (Alroy 1999; Roy 2001; Barnosky et al. 2004). In South America, 50 genera (83%) vanished during the arrival and spread of humans about 12,900 to 10,000 years ago. In Eurasia (excluding southern Asia), 9 genera out of 25 (36%) became extinct during two pulses (45,000–20,000 years ago, 12,000–9000 years ago). These extinction pulses correlate also with the spread and then the population increase in modern humans (Barnosky et al. 2004). In Africa, the losses were relatively mild with only 8 genera (18%). When considering only mammals > 1000 kg, the differences between continents are even more marked. In North America, all four genera were lost, Eurasia saw the demise of four out of five, but in Africa, no such genus went extinct (Roy 2001).

Many phases of human colonization were coeval with marked climate changes. Because earlier, similar climate changes were not accompanied by marked extinctions, hunting by humans (overkill) was proposed as the main mechanism responsible for the extinctions (Martin 1984), either by heavy and selective hunting (“Blitzkrieg”) or through habitat fragmentation, nonselective hunting and the introduction of exotic species (“Sitzkrieg”). The current consensus picture for megafaunal extinctions on the continents is that extinctions were most pronounced where a rapid spread and increase in H. sapiens populations coincided with marked climatic shifts (Burney and Flannery 2005). But it was not primarily large sized but rather slow breeding species (this is, in part, correlated with large body size), which were at the highest risk of extinction (Johnson 2002). Even if the proportion of deaths caused by humans was low at any one time, slow breeding megafaunal mammals were driven to extinction.

Yet the story is not over. Since the age of colonialization, the exploitation of nature has reached a new level, and the fate of the dodo (Raphus cucullatus) is just one very sad and telling example of human impact. This flightless bird was discovered in 1598 on Mauritius after Portuguese sailors first reached this island in 1507, and became extinct by 1690, not only by hunting but also by the introduction of domestic species, such as goats, pigs, and rats, which devoured the eggs and the young. The list goes on, and even extremely common species like the passenger pigeon in North America (Ectopistes migratorius) have proven to be no match for intensive hunting humans. Starting during the epoch of industrialization, exploiting nature and destroying natural habitats has proceeded...
at an ever increasing rate (Wilson 1994). There is no question that current extinction rates for plants and animals have reached a level perhaps 2–3 magnitudes above background rates (Nott et al. 1995; Pimm et al. 1995; Ricketts et al. 2005). Scaling the available estimates of current extinctions up to a magnitude where we can compare them to past Phanerozoic mass extinctions reveals that if species losses continue at the present rate, 96% of the species will be extinct within just a few hundred years (May et al. 1995; Sepkoski 1997; “Sixth” extinction). This is the maximum estimate for species losses during the most severe of all the Phanerozoic mass extinctions, the one that occurred at the Permian–Triassic boundary! Yet although the end-Permian mass extinction is no longer seen as a crisis spanning millions of years, current estimates are still on the order of a hundred thousand years. There is even further concern. As several episodes of mass extinctions have shown, even a moderate increase in global temperatures of a few degrees, if it happened fast enough, has proven fatal for life on the entire planet.

16.6 Conclusions

After more than 30 years of intensive research, the picture of life’s diversification on earth has attained an unprecedented level of accuracy. Much progress has been made in Precambrian research in the study of the oldest fossils and their environment. Major input into studies of radiations and extinctions came from geochronology, which provided a new, accurate time scale. Another important development was the establishment of a new animal systematic, and considerable advances were made in the dating of lineage splitting.

The quintessence of diversity studies, the global diversity trajectories for the Phanerozoic, have been refined in many ways, but the classical, iconographic figure of Sepkoski for the marine animals has changed little over the years. Yet many of the “dogmatic” concepts are reevaluated, a new data base is being developed, and methodologic problems receive considerable attention. What is needed for further refinements are especially studies at the local level that take into account ecological data and ultimately can be assembled into a new global picture.

Mass extinctions have received a tremendous amount of attention since 1980, and for most of them a consensus picture exists today. Emerging is that each extinction event had its own signature, and a common cause has not been found. Hot topics at the moment are the diversification of the metazoans, the Cambrian and Ordovician radiations, and here we will most likely see much progress in the next few years.
Acknowledgments

Various people at the Natural History Museum in Basel helped in stimulating discussions to improve this chapter. The chapter was critically read by W. Henke and I. Tattersall.

References


Berry WBN, Boucot AJ (1973) Glacio-eustatic control of Late Ordovician-Early Silurian platform sedimentation and faunal changes. Geol Soc Am Bull 84: 275–284


Bottjer DJ, Ausich WI (1986) Phanerozoic development of tiering in soft-substrata
suspension-feeding communities. Paleobiology 12: 400–420


Copper P (1986) Frasnian/Famennian mass extinction and cold-water oceans. Geology 14: 835–839


DeConto RM, Pollard D (2003) Rapid Cenozoic glaciation of Antarctica induced by
Patterns of diversification and extinction

declining atmospheric CO₂. Nature 421: 245–249


Hallam A (2002) How catastrophic was the end-Triassic extinction? Lethaia 35: 147–157


of the recent on the history of marine diversity. Science 300: 1133–1135


Kump LR, Pavlov A, Arthur MA (2005) Massive release of hydrogen sulfide to the surface...
ocean and atmosphere during intervals of oceanic anoxia. Geology 33: 397–400
Martin RE (1996) Secular increase in nutrient levels through the Phanerozoic:
Implications for productivity, biomass and diversity of the marine biosphere. Palaios 11: 209–219


Patterns of diversification and extinction


Sepkoski JJ Jr. (1979) A kinetic model of Phanerozoic taxonomic diversity. II. Early Phanerozoic families and multiple equilibria. Paleobiology 5: 222–251


Wade M (1972a) Dickinsonia: Polychaete worms from the late Precambrian Ediacara fauna, South Australia. Memoirs Qld Museum 16: 171–190
Wade M (1972b) Hydrozoa and scyphozoa and other medusoids from the Precambrian Ediacara fauna, South Australia. Palaeontology 15: 197–225
17 Paleoecology: An Adequate Window on the Past?

Thorolf Hardt · Britta Hardt · Peter R. Menke

Abstract

Starting from Ernst Haeckel’s famous definition of ecology, our review considers the premises and the meaning of paleoecological research. Unlike current ecology, paleoecology has to pay more attention when dealing with “facts”: the concept of uniformitarianism is presented and demonstrates the importance of philosophical constructs for scientific work. The middle-range theory attempts to filter out false conclusions. Abiotic factors have had a strong influence on adaptive evolution; volcanism, tectonism, and climate are exemplified. Subsequently we discuss the biotic viewpoint with regard to aspects of fossil findings and in this context taphonomy, stratigraphic research, and interactions between floral and faunal environment. In a synthesis, we present three cross sections in human evolution in different time horizons (early–middle–late) to exemplify the multidisciplinarity of paleoecology and present some key events that probably altered the direction of radiations among hominids. Obviously, human evolution is not a special kind of evolution; it follows strictly the rules of evolutionary biology and hence depends undoubtedly on environmental influences.

17.1 Introducing (paleo)ecology

In a strict sense, paleoecologists try to detect all processes which have affected a fossil organism antemortem (Behrensmeyer 1992). Postmortem events are analyzed by the methods of taphonomy, and by studying diagenesis (Andrews 1992; Lyman 1996), we have to realize that these two fields are inseparably linked in dealing with fossils and the reconstruction of their environments. Here, we illustrate briefly what (paleo-) ecology is and then we discuss actualism (or uniformitarianism) as an essential aspect of paleoecology.
17.1.1 Current ecology versus paleoecology

Haeckel (1866), who introduced the term “Ecology” (German: Oecologie, nowadays Ökologie) stated:

- Ecology is a part of science that deals with the relationships between organisms and their surrounding environment, wherein we can place all conditions of being” in the broadest sense. These are of partly organic, partly inorganic nature; and both are of utmost importance for the form of organisms, because they force it to adapt to them (translated and shortened by the authors).¹

Haeckel included the following aspects in his “conditions of being”:

1. **Abiotic factors** (physical, chemical, climatic, electricity-conditions, inorganic food, composition of water and soil).
2. **Biotic factors** (all relationships between organisms).

More recently, modern ecology has been defined as “the study of the relations between organisms and the totality of the physical and biological factors affecting them or influenced by them” (Pianka 1983 p 3). Among its most interesting branches are the distribution and frequency of organisms/populations/communities, hence natality, mortality, and migration (Begon et al. 1998). It makes sense to divide ecological research into three subsections (Table 17.1).

All of these definitions are crucial when talking about ecology; however for several reasons, it is not that easy to simply insert the prefix “paleo” for the study of ancient ecologies (Andrews 1992).

There are not only strong connections between paleoecology and recent ecology but also some basic differences (Table 17.2).

---

Paleoecological research is aimed mainly at analyzing long-term trends because when dealing with fossils short-term processes (e.g., successions, micro-evolution) are not recognizable, or recognizable only with difficulty. Frequently, paleoecological analyses focus on the development of communities in certain paleoenvironments over time spans of millions of years (Stanley 1994). Despite the fact that there is a conspicuous nexus between organismic paleoecology and geological orientated facies observations, it is preferable in practice to separate the two, since both disciplines are in themselves multilayered and sophisticated networks (Etter 1994). Contemporary paleoecological questions include:

Are similar as well as different morphological characters linked to adaptations in the same kind of habitat?
Are there interdependencies between paleomilieu and life cycles/life-history parameters/population densities?
How is distribution and diversity of life on earth dependent on ecological requirements?
Can extinction events be traced back to major changes in the environmental conditions, and if so, to which?

All of these questions require a high degree of multidisciplinarity and cannot be solved only in terms of one-factor analyses. They also have to encompass multiple factors that build ecosystems like the one covering our planet.

After a short look at the approach of actualism/uniformitarianism, we will take the previous quotation by Haeckel as a guidepost for the following sections. After touching on abiotic factors we proceed to the biotic ones, and finally merge them in a third step that we designate as synthesis. In this we try to focus on hominid or near-hominid evolutionary perspectives and leave out important paleoecological aspects that refer to marine ecosystems.

17.1.1.1 “The present is a key to the past”: A valid premise?

One of the most important concepts in the geosciences is uniformitarianism, a principle introduced by the Scottish physician and geologist James Hutton (1726–1797). Originally, this was framed as an antagonist to the idea that catastrophic phenomena might have formed Earth’s surface. It was the establishment of the idea that the laws of nature stay constant that made geology a mature part of the scientific endeavor. This philosophical approach is axiomatic in physics, but it was Sir Charles Lyell, “Darwin’s guru and intellectual father figure” (Gould 1994 p 6764), who placed this idea before a broader scientific community. The application of this principle inspired Darwin, although Gould (1994) warned against the pure extrapolationism of Darwin’s uniformitarian perspective. Gould’s (1994 p 6768) attractive musings on paleontology’s meaning are more important than ever. The simple-mindedness of universal reduction to lower levels must be abandoned: “Our evolutionary world is a hierarchy of levels, each of legitimacy and irreducible worth.” Riedl (1981) also emphasized the limitations of each methodological level.

Based on Gould (1984), Etter (1994) reduced the philosophy of actualism to four principles:

1. Laws of nature do not change on Earth in time and space.
2. Processes that influenced geological phenomena in the past occur in the same manner now.
3. The speed of geological and biological processes does not change.
4. In the past the same materials and the same conditions existed.

The first two of these are methodological assumptions that are necessary to conduct inductivist research. Points three and four are less definite. As Etter (1994) points out, the adoption of all four premises may be limited due to (1) constraints of observing the past and (2) the strong fluctuations of geological phenomena which occurred long ago (e.g., the circumstances of the dinosaurs’ extinction).

Dodd and Stanton’s (1990) taxonomic uniformitarianism is a derivative of substantive uniformitarianism and an attempt to reconstruct ecological niches by assuming that the environment of a fossil will have been identical with that of the nearest extant relative. But it is obvious that reliable conclusions can only be made in this way if a fossil is a member of an extant species. This can virtually never be the case except for Pleistocene and Holocene deposits (Etter 1994), and even then there are also theoretical problems, one of which is the fragmentary knowledge we have of former ecological niches. An absence of enemies could, for example, have caused an expansion of the niche. Thus Etter (1994) admits that taxonomic uniformitarianism should be limited to fossils with extant representatives or relatives. Yet the method can be improved by investigating groups of species within a taxonomic group rather than just one.

Tattersall (1998) reviewed such provocative ideas of evolutionary biologists as Dawkins’ “Selfish Gene” or Eldridge and Gould’s “Punctuated Equilibrium” and also relativized the meaning of adaptationism (p 95: “…organisms may not be as exquisitely fine-tuned to their environments …”) in our evolutionary thinking. Therefore, we should not forget that there exists a possibility of overstraining positivism. Actio-reactio thinking is of course essential, and the only way of doing science in general; but the danger of story-telling in paleoecology is acute. We are at risk of delivering explanations for adaptations which could be true but actually might not be (see later). It is questionable if the equation: \( \text{Bauplan:environment} = 1:1 \) is satisfied.

Foley (1978) linked the past and the present by summarizing Binford’s (1977, 1981) middle-range theory: this is actually an assemblage of theories linking observable aspects of the past to processes operating today, permitting inferences about the conditions of geological and biological systems in the past. Such observations must be linked to observable past phenomena (i.e., an anatomical element) (Figure 17.1). A connecting piece between a theory (deductions and predictions from a basis of axioms) and the data is the model (which describes and predicts the variation and structure of phenomena derived from theoretical principles).
Foley (1987) also points out that there are different levels of model building, depending on behavioral (predispositional) and geological (postdepositional) factors. Furthermore, identical records might be produced by different behaviors (Hill 1984; Foley 1987), so the overlap of different models has to be taken into consideration. Four pathways are very useful for testing a theory in order to reduce the pure reflection of contemporary ideas: (1) a careful analysis of each link in an inferential chain, (2) like Binford (1981) not following the broad trend but rather looking for small details that might appear comparatively insignificant, (3) the isolation of processes that may result in similar results and assigning (theoretically/experimentally) the probabilities of various outcomes, and finally, (4) the comparative biological approach, which investigates patterns of interspecific and intraspecific variation that can be used to correlate biological variables (Figure 17.2).

17.1.1.2 Ecology and cladistics

The a priori assumption that a particular character (or character state) is the result of a particular evolutionary process may lead to far greater problems of circularity than does the incorporation of these characters into a global estimate of evolutionary history (Luckow and Bruneau 1997 p 150). These authors infer that the exclusion of ecological information from a phylogenetic analysis when testing ecological hypotheses is not only unnecessary but also “undesirable.” Luckow and Bruneau (1997) justified their arguments by concluding that character exclusion...
would partition the data in an arbitrary way and that discrete homology statements would get lost.

17.2 Abiotic factors

Throughout natural history, abiotic factors have had a strong influence on evolution, adaptive or otherwise. Environmental change has shaped habitats and the evolutionary “fates” of living systems. One fundamental aim of evolutionary biology is to understand and reconstruct the interaction of environmental abilities and the survival of organisms. Wegener (2005) wrote: “Only by summing up all fields of the geosciences may we hope to find out the “truth,” this means determining a picture that integrates all known facts in the best order and which seems to represent the utmost probability; even then we have to be aware that each discovery, regardless of which scientific field it emanates from, may modify the result”² (Translated by the authors). Early acceptance of this ignored genius’

² “Nur durch Zusammenfassung aller Geo-Wissenschaften dürfen wir hoffen, die ‘Wahrheit’ zu ermitteln, d.h. dasjenige Bild zu finden, das die Gesamtheit der bekannten Tatsachen in der besten Ordnung darstellt und deshalb den Anspruch auf größte Wahrscheinlichkeit hat; und auch dann müssen wir ständig darauf gefasst sein, dass jede neue Entdeckung, aus welcher Wissenschaft immer sie hervorgehen möge, das Ergebnis modifizieren kann.” (Wegener 2005 p 222)
appeal for an adequate scientific network might have accelerated the unraveling of Earth’s physical secrets.

17.2.1 Geological influences

17.2.1.1 Volcanism

In the truest sense of the word, volcanoes put pressure on their environment. Usually eruptions have a localized but strong influence. However, tephra (ejecta, e.g., ash or pumice) and lava do have the potential to modify water and soil chemistries, eliminating or modifying habitats over wide areas, or supporting the conservation of important fossil footprints.

Including sea bottom and continents, about 80% of the Earth’s surface has been produced by ascending melted rock. The new material cools down and hardens. This activity is primarily constrained to plate boundaries. Volcanic material creates a useful (blurred) perspective concerning Earth’s interior. Rocks also deliver nutriments, chemical resources, and minerals (Press and Siever 1995).

Feibel (1999) investigated the eruptive activity in the Turkana area of northern Kenya and concluded that the higher assumed rate (38 events per 240,000 years) indicates one significant eruption per 6,300 years. Massive eruptions can also cause global impacts since they glut the atmosphere with aerosols. Rampino and Shelf (1992) discussed the controversial idea of the Mount Toba-explosion 70,000 years ago, which might have produced a sharp cooling and consequently a shift to glacial conditions.

17.2.1.2 Tectonic aspects

Faulting within local sedimentary basins, continental plate movement, and the highland-formation through uplift may greatly affect terrestrial habitats.

Between 16 and 12 Ma, the Eurasian and the Afro-Arabian continental plates moved and allowed an exchange of the flora and fauna: e.g., an exodus of early African apes (Potts 2003). Between 7.0 and 5.0 Ma, the Atlantic Ocean and the western end of the Mediterranean were temporarily separated as Africa drifted northward. Here, a periodic drying and flooding of the Mediterranean basin occurred and huge salinity deposits were built up. Western Eurasia dried due to an evaporation and salinization phenomenon. Between 4.5 and 3.0 Ma, the Isthmus of Panama was formed by the contact of continental plates and
### Table 17.3
The Ethiopian Rift System and some important paleoanthropological sites

<table>
<thead>
<tr>
<th>Areas with paleoanthropological sites</th>
<th>Location within the rift</th>
<th>Sedimentary environments</th>
<th>Age of tuffs</th>
<th>Number of tuff beds/Ma (preserved)</th>
<th>Tuffs; percentage of stratigraphic section</th>
<th>Percentage of tuffaceous sediments in stratigraphic column</th>
<th>Hominid species</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shungura Usno Fejej</td>
<td>Omo rift</td>
<td>Lacustrine, fluvial, overbank</td>
<td>4.1–1.39 Ma</td>
<td>13</td>
<td>9</td>
<td>Estimated at 70</td>
<td>A. afarensis, A. aethiopicus, A. boisei, H. habilis, H. erectus</td>
</tr>
<tr>
<td>Konso-Gardula</td>
<td>Southern MER (Main Ethiopian Rift)</td>
<td>Fluvial</td>
<td>2.0–1.35 Ma</td>
<td>~4</td>
<td>5</td>
<td>95</td>
<td>H. erectus</td>
</tr>
<tr>
<td>Gademota (Middle MSA)</td>
<td>Central MER</td>
<td>Collovium, paleosol</td>
<td>35–1.81 ka</td>
<td>3</td>
<td>n.d.</td>
<td>n.d.</td>
<td>H. sapiens</td>
</tr>
<tr>
<td>Gadeb</td>
<td>Eastern rift shoulder, Central MER</td>
<td>Lacustrine, fluvial</td>
<td>2.51–ca. 0.7 Ma</td>
<td>&gt;2</td>
<td>~12</td>
<td>Estimated at 50</td>
<td>H. erectus</td>
</tr>
<tr>
<td>Melka Kunture</td>
<td>Central MER</td>
<td>Fluvial</td>
<td>&lt; or = 1.5 Ma</td>
<td>5</td>
<td>n.d.</td>
<td>n.d.</td>
<td>H. erectus</td>
</tr>
<tr>
<td>Kesem-Kebena (K-K6)</td>
<td>Northern MER</td>
<td>Lacustrine, fluvial, lacustrine</td>
<td>1.04–1.0 Ma</td>
<td>3</td>
<td>38</td>
<td>62</td>
<td>H. erectus</td>
</tr>
<tr>
<td>Middle Awash</td>
<td>Southern Afar</td>
<td>Fluvial, overbank, lacustrine</td>
<td>4.38–4.29 Ma</td>
<td>&gt;20</td>
<td>5</td>
<td>40</td>
<td>A. ramidus, H. erectus, A. garhi</td>
</tr>
<tr>
<td>Hadar</td>
<td>West-central Afar</td>
<td>Fluvial, overbank, lacustrine</td>
<td>3.4–3.18 Ma</td>
<td>5</td>
<td>&lt;5</td>
<td>Estimated at 40</td>
<td>A. afarensis, H. erectus</td>
</tr>
<tr>
<td>Gona</td>
<td>West-central Afar</td>
<td>Lacustrine, deltaic, fluvial</td>
<td>2.94–&lt;2.52 Ma</td>
<td>&gt;4</td>
<td>&lt;5</td>
<td>Estimated at 40</td>
<td>None</td>
</tr>
</tbody>
</table>

Modified from WoldeGabriel et al. 2000.
the following uplift (Stanley 1995). One result was the strengthening of the conveyor belt of Atlantic Ocean currents: the North Atlantic was “watered” by the warm Gulf Stream. The initiation of ice ages in the Northern Hemisphere during the Late Pliocene can be explained by the development, over 36 Ma, of isolation of the Arctic Ocean caused by a high-salinity, warm, sinking, and returning current (Stanley 1994).

The rift valley formation in eastern Africa, the uplift of western North America and the Tibetan Plateau represent examples of drying and cooling effects. Models of general circulation show that when major air currents are divided by elevated plateaus, the air is altered due to winter cooling and summer heating causing high- and low-pressure areas to form over landmasses far from the plateaus (Ruddiman and Kutzbach 1989). One consequence is a greater seasonal variation and also the creation of seasonal monsoons. The overall effect is a cooler and drier global climate (Ruddiman et al. 1989). Even geochemical weathering on rocks can be traced back to uplifts: enhanced monsoons, steepening river gradients, and faster erosional rates increase weathering. Raymo and Ruddiman (1992) suggested that global cooling might result from the deposition of carbon from the highlands into the ocean as well as from weathering responses. A consequence would be a reduction in carbon dioxide and a removal from the atmosphere. The heat-retaining function—carbon dioxide is a greenhouse gas—would disappear. Another effect occurs on the leeward side of the uplifted region: upland depletion of air moisture and precipitation cause rainshadow drying, a phenomenon partly responsible for the aridification of the African rift valleys compared with its western counterpart.

Another possibility is that the shapes of sedimentary basins are modified by the local impact of earthquakes. Availability of water to the local biota is obviously a very important factor, so the changes of fluvial systems by intrabasin faulting mechanisms are also a matter of paleoecological discussion.

Eastern Africa is a superb example to illustrate intracontinental rifting phenomena. During the Cenozoic, a system of continental rifts developed from the Red Sea and the Gulf of Aden in a southward direction from the Afar-region of Ethiopia, which is a triple junction where three developing tectonic plates come together (Figs. 17.3 and 17.4; Table 17.3). This phenomenon is widespread. A former triple junction existed where Africa and Northern and Southern America came together. The Amazon, the Mississippi, and the Niger represent rivers that run through “failed rifts.” Afar is characterized by three spreading zones and represents a small portion of oceanic crust that has been uplifted and is now part of the continent (Stanley 1994).
Figure 17.3
African Great Rift Valley system (a) and Rifting process (b) with accentuation of the Afar triangle and the Turkana basin (c)
17.2.2 Climatic influence

17.2.2.1 Cycling planet Earth

The Milankovitch cycles (Alverson Volume 1 Chapter 12) are a consequence of the changing gravitational pull of other planets. Variation in solar heating, which is related to astronomical cycles altering orientation and shape of Earth’s orbit around the Sun, is attributed to these periodic oscillations. The solar radiation, its distribution and strength, changes over the globe in periods of 19, 23, 41, and 100 thousand years (Hewitt 2003). How can these periods be detected? Oxygen isotopes are documented in the deep-sea record and represent measurements of the global ice volume, ocean temperature, and evaporation. Around 3.0–2.5 Ma, the amplitude and frequency of $\delta^{18}$O oscillation changed significantly. Northern Hemisphere glaciation and greater aridity in the tropics correspond to this pattern. Between 900 and 600 ka, the effects of 100,000-year-long cycles of glacial forming and interglacial warming came into play (Hewitt 2003).
The 23,000-year cycle of Earth’s rotational axis intensifies African monsoons, and the millennial-scale instability causing North Atlantic iceberg discharge involving huge masses of ice rafting, melting, and a decrease of ocean salinity was also important here. Taylor et al. (1993) and Bond et al. (1997) traced this phenomenon back to decade- to century-scale fluctuations between near-glacial conditions and interglacial warmth. Raymo et al. (1998) and McManus et al. (1999) linked these conditions to the Late Pleistocene and Early Holocene.

Well-defined cycles are not necessarily well separated in terms of amplifying and buffer mechanisms. Their interactions show complex patterns, as Clemens et al. (1996) and Paytan et al. (1996) have demonstrated. Furthermore, Potts (2003 p 365) stresses that “major environmental shifts occurred episodically throughout the Quaternary and did not necessarily coincide exactly with maximum changes in temperature, moisture, or any other single factor.” Lister and Rawson (2003) called attention to the rise and fall of sea level, one major effect of the climatic ups and downs: over the past 600,000 years fluctuations of up to 120 m are documented. Inundations as well as the exposure of continental shelves created new barriers—or pathways.

17.2.2.2 Our planet’s oceans

Earth’s climates were much wetter and warmer before Middle Eocene times, and deciduous and evergreen forests dominated the natural scenery. Antarctica was full of temperate rain forests and Arctic land masses were distributed with trees (Askin 1992; Denton 1999). A strong seasonality and a general lack of aridity reduced the occurrence of grasslands and deserts. Sea ice and glaciers were limited concerning their volume expansion (Ruddiman and Kutzbach 1991). The world’s oceans were about 10°C warmer at depth than today. Furthermore, the atmospheric temperature gradient—from the equator to the pole—was much smaller than that of today (Savin et al. 1975; Shackleton and Boersma 1981).

This changes when we look at the recent situation. Since the Middle Eocene, global climate has cooled. Shackleton (1995) discussed the essential global paleoclimatic change at 2.95–2.52 Myr, and the beginning of the oscillating ice ages. Over the past 0.95 Myr, each glacial maximum was on average 5°C cooler than today and the planet was drier, e.g., much of northern Europe was covered by tundra and Africa was drier (Denton and Hughes 1981; Denton 1999).
17.2.2.3 Another unique continent: Africa

deMenocal (2004) gave a very useful overview of the African climate: as Figure 17.5 demonstrates, the combination of the seasonal migration of the intertropical convergence zone and the African monsoon causes a highly seasonal range of North African rainfall. In the boreal summer, heat over the North African land surface, centered near 20°N, draws moist maritime air from the equatorial Atlantic into western and central subtropical Africa. Accordingly, woodland and grassland savannahs flourish (Hastenrath 1985; Harris 1980).

Based partly on topographic rainshadow effects, summer rainfall in East Africa is very variable. It is also related to the westerly airstream of the African monsoon (Nicholson 1993). East Africa's subtropical rivers, such as the Omo and the Nile, are drained by the summer monsoonal runoff via the capture of moisture by Ethiopian and Kenyan Highlands. On the other hand, relative to adjacent oceans, Asian and African landmasses become cooler and a reversion of the atmospheric circulation comes about (deMenocal 2004).

Prospero and Nees (1986) showed that the changes of subtropical African summer rainfall are closely tied to West African dust export to the Atlantic. Additionally, these modifications have been linked by Giannini et al. (2003) to anomalies of the sea-surface temperature. Figure 17.5b demonstrates that Indian monsoon surface winds are interconnected with summer dust plumes of Arabia and northeast Africa (Nair et al. 1989). These winds carry mineral dust to the Arabian Sea and the Gulf of Aden. The study of mineralogical and sedimentological data reveals that wind-borne detritus “from these sources comprises the dominant source of terrigenous sediment to the eastern equatorial Atlantic and the Arabian Sea” (deMenocal 2004 p 7).

Marine paleoclimatic records The last (ca.) 5 Myr, during the Late Neogene, showed progressive step-like increases in African aridity and periodical arid–humid climate cycles. This conclusion can be drawn from marine sediments accumulating off the western and eastern margins of the subtropical North African region. The isolation of the Atlantic basin via the Isthmus of Panama and the following gradual onset of high-latitude glacial cycles at 3.2–2.6 Ma seem to have influenced African climate variation patterns (Haug et al. 2001). The onset of glacial ice rafting and modest 41-kyr glacial cycles after 2.8 Myr caused Plio-Pleistocene cooling at high latitudes. Another step was a shift toward cooler conditions and, after ca. 1.6 Myr, higher-amplitude 41-kyr cycles after 1.2–0.8 Myr (Shackleton et al. 1984).
Figure 17.5
(a) Regional map of Northern Africa vegetation zones, locations of DSDP and ODP drill sites (filled circles), and locations of selected African mammal fossil localities (open diamonds).
(b) Boreal summer (August) surface wind stress (unit vector = 1 dyne/cm²), intertropical convergence zone (ITCZ, heavy dashed line) location, and boundaries of seasonal tropospheric dust plumes off NW Africa NE Africa/Arabia. Dust plume contours were derived from haze frequency data. (c) Boreal winter (January) surface wind stress (unit vector = 1 dyne/cm²), ITCZ location, and boundary of the seasonal tropospheric dust plume off NW Africa (reprinted from Earth Planet Science Letters, Vol. 220, de Menocal, P.B., African climate change and faunal evolution during Plio-Pleistocene, P. 3–24, with permission from Elsevier.)
Variability of subtropical African paleoclimate: deMenocal (2004) summarized the patterns of marine sediment records of Plio-Pleistocene eolian export from West and East Africa:

- The variability of orbital-scale African climate variability persisted throughout the entire interval (and in some cases extending into the Miocene and Oligocene)
- The onset and amplification of high-latitude glacial cycles was closely linked to the onset of large-amplitude African aridity cycles
- A gradually increasing after 2.8 Myr of eolian concentration and supply (flux)
- At 2.8 (±0.2), 1.7 (±0.1), and 1.0 (±0.2) Myr, step-like shifts in the amplitude and period of eolian variability
- There is evidence for $10^4$ to $10^5$ year “packets” of high- and low-amplitude paleoclimatic variability which were paced by orbital eccentricity.

deMenocal (2004 p 8) described the marine record as “a succession of wet-dry cycles with a long-term shift toward drier conditions, punctuated by step-like shifts in characteristic periodicity and amplitude.” de Menocal (1995) interpreted subtropical African climate prior to 2.8 Myr varying at the 23–19-kyr period, mainly as a result of Asian monsoonal variability. At 2.8 (±0.2) Ma, there was a shift toward climate variation periods longer than 41 kyr and after 1.7 (±0.1) Myr the cycles lengthened again, and an eolian variability-shift toward longer and larger-amplitude 100-kyr cycles after 1.0 (±0.2) Myr occurred. The onset and growth of high-latitude ice sheets and cooling of the subpolar oceans were synchronous with these shifts in the African eolian variability (Shackleton et al. 1990), and there was a coupling between high- and low-latitude climates after the glaciation-onset near 2.8 Ma (deMenocal 1995).

Dupont and Leroy (1995) showed that a pollen record from Site 658 documents the phenomenon of greater variability and progressively xericasive vegetation after ca. 3.0 Mya and concluded that a shift toward a drier and cooler African climate occurred during glacial maxima. The pollen record correlates with oxygen isotopes indicating that large global ice volume and deep-sea temperatures correspond to aridity. These authors conclude that a comparison of short-term fluctuations of periods before and after 2.5 Myr demonstrates “that obliquity forcing of northwestern African climate started with the first large glaciations in the Northern Hemisphere” (p 297).

Which phenomenon is the fundamental impulse generator for African climate variability? The Plio-Pleistocene succession of sapropel layers in the Mediterranean Sea could be interpreted as a consequence of orbital precession
(Hilgen 1991). Enhanced monsoonal and Nile river runoff led to increased Mediterranean stratification and reduced ventilation of the deep eastern basins (Rossignol-Strick 1985). During these humid periods, organic-rich layers were deposited. However, another stimulating factor could be a covariation of African climate with the high-latitude climate cycles at the 41- and 100-kyr periodicities, which is what marine sediments actually suggest.

deMenocal (1995, 2004 p 10) tried to reconcile the two different points of view “by acknowledging that precession was the fundamental driver of African monsoonal climate throughout the late Neogene, but that high-latitude glacial cooling and drying effects were superimposed on this signal only after 2.8 million years.”

17.2.3 Physical (paleo)geography or the beauty of mosaics

The African continent extends virtually equidistant into both the Southern and the Northern Hemispheres. It includes about 20% of the world’s land surface and stretches 8,000 km from north to south. The configuration of the bordering oceans and landmasses has remained almost the same from the Early Pliocene up until today (O’Brien and Peters 1999).

Based on the pioneering work of Lobeck (1946), O’Brien and Peters (1999) subdivided Africa into different physiographic regions (Figure 17.6). In Low Africa, during most of the Pliocene, all of the interior basins may still have lacked outlets to the sea, a condition not existing today. Prerift Africa’s (Miocene–Pliocene) interior basins and associated drainage systems are nowadays etched by old deltas, strandlines, old terraces, and alluvial deposits. This might indicate (at least) seasonally expansive internal lakes, e.g., the Paleolake Congo. This was probably a perennial water body up to at least the Late Pliocene, and its eastern catchment extended to the High Interior Plateau and the volcanic highlands of the Eastern Rift Belt (O’Brien and Peters 1999).

In the Oligocene, High Africa was tectonically driven by the African Swell, which was active again in the Late Pliocene and Pleistocene. Different effects have influenced southern and eastern Africa: during the Late Pliocene, local environmental and increasing subregional fragmentation took place. During the Early as opposed to the Late Pliocene—up to 3.0 Myr—the rift grabens associated with doming were shallow and also at higher elevations than now (Brown 1995). The Ethiopian Massif was probably lower than now, while the High Eastern Interior Plateau may have been higher (Feibel 1999). Between 3.0 and 2.0 Ma, an uplift of rift shoulders by 1,000–1,500 m and a concomitant major subsidence of rift
grabens occurred: (1) first in the Eastern Rift Belt areas and in the Afar-region and later (2) in the Western Rift Belt, where some of the uplifted flanks rose to 4,000 and more meters above sea level (Partridge et al. 1995a, b). The eastern drainage perimeter for the Congo Basin was established by the Escarpment Mountains of the Western Rift Belt. They also caused a diversion of the westward drainage from the Eastern Rift Belt and the Interior Plateau, into the Sudd. The Rift Valley Domain of the East African Highlands was particularly perforated by active volcanoes (O’Brien and Peters 1999).

Africa’s Southern Platform was formed by subsidence events resulting in local fragmentation of the environment (exceptions: areas immediately adjacent to subsided or uplifted margins) and by broad-scale regional uplift. The Southern Platform was apparently lower in elevation than nowadays by some 200 m in the southern escarpment mountains and by as much as 1,000 m in the extreme southeast. The Transvaal, however, was somewhat higher than today (maybe about 400 m). Furthermore, the paleolakes of the Southern Platform are of special interest for hominid ecogeography (e.g., Makarikari). The grabens of the
Western and Eastern Rift belts in the East African Highlands contained intermittent rivers and lakes. Lake Malawi, however, obtained its present shape only at the end of the Pliocene (Partridge et al. 1995b; O’Brien and Peters 1999).

### 17.3 Biotic factors

Together, abiotic factors build a framework for living spaces that frequently harbor more than one population (Ziegler 1992). These spaces are called *biotopes*, with their enclosed organismic communities characterized as *biocenoses*. These often include highly adapted and specialized species in numerous ecological niches; together they build so-called ecosystems or *biomes*. The goal of paleoecologists is to reconstruct these systems or paleoenvironments based on organic and inorganic remains.

In the following section, we concentrate mainly on fossils, their interpretation, and the surrounding factors that influence them and thus have to be analyzed when reconstructing paleoenvironments. We touch on questions of *post mortem* processes, stratigraphic research, and coevolution. Due to the fact that we spotlight hominid and accordingly terrestrial evolution, we omit an extensive debate on plant fossilization and exclude aquatic ecosystems.

#### 17.3.1 Fossils

A fossil is any remain or trace of any organism from all past periods. They are the key elements of paleontology and consequently of paleoanthropology and paleoecology and act as *containers* or *archives* that preserve information over a long period. Scientists who deal with fossils have to find the right *tools* to open and unravel the secrets lying within. We try to illustrate some of these “tools” and enlighten several approaches of paleoecological research.

**17.3.1.1 What happened *antemortem* and what *postmortem*?**

It is unlikely that an organism will be preserved through time after death. Less than 1% of all organisms are handed down to us (Ziegler 1992; deMenocal 2004), especially in terrestrial and tropical conditions, as we see in the case of hominoids (Carroll 1988; Martin 1990; Stanley 1994; Andrews 1990). Due to this, the density of relevant paleoanthropological fossils is very low—approximately one fossil per one hundred generations—and there are a lot of gaps that effect the phylogenetic and ecological reconstructions scientists strive for (see later). Besides this, most fossils are fragmentary or incomplete and require reconstruction by means of all
available methods, now including 3D-reconstruction using CT, MRT, or surface scanning (Ulhaas Volume 1 Chapter 27). Additionally, we have to consider the limited erosion of noteworthy ancient fossil sites on the African continent, since most areas are covered with tropical forest, as well as the political and financial circumstances that impede excavations on the other hand.

Every fossil contains a great deal of information on the evolutionary history (phylogeny), physical organization (morphology), and lifestyle (ecology) of the populations the organism once belonged to. Under ideal circumstances, it is possible to extract most of this information by careful analyses of the taphonomic and diagentic processes involved. Taphonomy, as a special field of paleontology, represents the description and causal analysis of all factors that influence an organism after death (necrology and biostratinomy) (Figure 17.7) and subsequently all processes of embedding and fossilization (Efremov 1940; Lyman 1996; Grupe Volume 1 Chapter 7). In contrast, diagenesis characterizes only the biological, physical, and chemical alteration of the mineralogical elements affecting fossil-bearing sediments and is part of taphonomy and the lowest grade of rock-metamorphism (Stanley 1994; Lyman 1996; Conroy 1997).

The possibility that an organism will be fossilized after death—its transition from the bio- to the lithosphere—depends on many environmental factors (Foley 1978; Henke and Rothe 1999). With few exceptions, the soft tissue undergoes the physiological processes of decomposition and decay in the first phase after death (necrology), and normally only hard materials like teeth, bones, shells, and rarely scales and horn are suitable for being mineralized and thus fossilized. The most important participants in this process are: calcium in the form of calcite and aragonite (CaCO₃), bone apatite (Ca₅(OH)(PO₄)₃), silicic acid—skeletal opal (SiO₂), chitin, cellulose, and scaffold proteins (spongine, ceratine).
17.3.1.2 Who are you? Where do you come from? What brought you there?

The only way to reconstruct ancient ecosystems is to examine the fossil record and elucidate the internal and external influences that led to the life forms we try to understand and explain. Following Etter (1994), there are three crucial requirements that have to be fulfilled when analyzing fossils:

1. Accurate determination and systematic classification of the collected specimen.
2. Putting all investigated profiles in a temporal and stratigraphical order as precisely as possible.
3. Understanding the ecological context and the specific adaptations of the organisms that enable it to live in particular environments.

In brief, we have to study fossils in context to understand the processes occurring in interrelated evolving systems (Figure 17.8).

17.3.1.3 Who?

Taxonomic classification is not only essential but also complicated and depends on whatever evolutionary theory, species concept, and taxonomic approach the
researcher prefers (Delson 1990; Foley 1978; Tattersall 1992, 1996; Wood 1992, 1996, Wood and Collard 1999; Wolpoff 1999; Sarmiento et al. 2002). All these questions go beyond the scope of this chapter and are discussed elsewhere in these volumes (Ohl Volume 1 Chapter 4). We need to overcome the subject–object problem (Stadler et al. 1977; Vogel 2000) when analyzing how we were “becoming human,” or even when assembling or disassembling new genera or species—not to mention new hypotheses or theories—as soon as possible after finding new hominid remains. In the words of Eric Delson: “The paleoanthropology community must look quite Pavlovian to outsiders—we all drool predictably every time a new fossil is discovered” (Delson 1997 p 445). Obviously, new fossil findings complicate the puzzle of primate and hominid evolution (Foley 1991; Tattersall 1992).

Since Darwin’s ideas about natural selection and his groundbreaking theory of evolution via selective forces were published in 1859, much has been learned about the processes that generate species and the debate is still in progress. Using the individual specimens that constitute the fossil record, we can indirectly scrutinize ancient organic systems at definite points in time (“Semaphoronts” sensu Hennig 1982). One affiliated evolutionary theory, “punctuationism,” (antonym: gradualism) postulates rapid evolutionary development at the nodes of the “Tree of Life,” and long periods (branches) with slow rates of evolutionary changes. Associated is a crucial problem of fossil research, for the chance of finding “node-fossils” (once called “missing links”) is significantly lower than of finding “branch-fossils.” Essentially, human evolution is strongly embedded within the framework of evolutionary biology and has to be seen as a chain of adaptive radiations (Foley 2002) and extinction events.

17.3.1.4 Where and when?

One of the most important factors here is the time dimension and thus the absolute and relative determination of age (dating). Stratigraphers subdivide sediment layers chronostratigraphically and put them in a hierarchic system (Table 17.4; Figure 17.9). Each unit within a stratigraphic system has been defined by means of a basal Global Standard Section and Point (GSSP) by the International Commission on Stratigraphy. Absolute dating, or chronometrics, is done by direct chemical (e.g., amino acid racemization) or physical (e.g., radiometric, thermoluminescence, magnetism) analyses of bone tissue, rocks, or fossil material in the facies of interest, and different techniques attain different time

---

3 Facies: the sum of all primary rock characteristics; incorporates inorganic (lithofacies) and organic (biofacies) elements.
Relative dating methods are tightly related to the theory of, or directly on, stratigraphy and the idea of marker fossils, primarily developed by Nicholas Steno, a Danish anatomist, and upgraded by William Smith (alias “strata Smith”), an eighteenth-century English engineer (Ziegler 1992; Stanley 1994; Rothe 2000). Their perceptions led to three principles of sedimentation: (1) younger deposits overlie the older ones from the bottom up, (2) in the initial state all layers are approximately horizontal, and (3) sediments extend laterally and can be parallelized over large areas (Rothe 2000).

### Table 17.4
Summary of the categories and unit-terms in stratigraphic classification

<table>
<thead>
<tr>
<th>Stratigraphic categories</th>
<th>Principal stratigraphic unit terms</th>
<th>Equivalent geochronological units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lithostratigraphic</td>
<td>Group</td>
<td>Note: If additional ranks are needed, prefixes “Sub” and “Super” may be used with unit-terms when appropriate, although restraint is recommended to avoid complicating the nomenclature unnecessarily</td>
</tr>
<tr>
<td></td>
<td>Formation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Member</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bed(s), Flow(s)</td>
<td></td>
</tr>
<tr>
<td>Unconformity-bounded Biostratigraphic</td>
<td>Synthem</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Biozones:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Range zones</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Interval zones</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lineage zones</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Assemblage zones</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Abundance zones</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Other kinds of biozones</td>
<td></td>
</tr>
<tr>
<td>Magnetostratigraphic polarity</td>
<td>Polarity zone</td>
<td></td>
</tr>
<tr>
<td>Chronostratigraphic</td>
<td>Eonothem</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Eratham</td>
<td></td>
</tr>
<tr>
<td></td>
<td>System</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Series</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stage</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Substage</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Chronozones)</td>
<td></td>
</tr>
<tr>
<td>Other (informal) stratigraphic categories (mineralogical, stable isotope, environmental seismic, etc.)</td>
<td>-zone (with approximate prefix)</td>
<td></td>
</tr>
</tbody>
</table>

Murphy and Salvador 1999.
A further important law was established by Johannes Walther, a German geologist and student of Ernst Haeckel, who based his inferences on the Swiss geologist Amanz Gressly’s fundamental contributions on stratigraphy in the eighteenth century and stated that the vertical succession of facies reflects lateral changes in past environments (Cross and Homewood 1997).

Stratigraphic research can be roughly divided in three parts that differ in materials and methods:

(a) *Lithostratigraphy* deals with the sequence and succession of sedimentary beds to yield diachronic zones; only a few “key beds” are nearly isochronic (Behrensmeyer 1992; Stanley 1994).

(b) *Biostratigraphy*, which is based on organismic evolution, shows overlapping zones by analyzing facies-specific marker or index fossils that can be deemed as isochronic (Stanley 1994). Age estimations become more precise, the more closely index fossils are related (Behrensmeyer 1992). The lifespan of a fossil species, referred to as its “zone,” ranges from the first appearance of that species until a subsequently following species replaces it. Biostratigraphic timescales are often bound to a specific area. Both approaches offer the possibility of correlating sediments and time horizons continuously in regional areas (Parastratigraphy) and worldwide (Orthostratigraphy).
Chronostratigraphy, rock layers (strata) are sequenced and classified by absolute dating methods.

Christian Leopold von Buch, an eighteenth-century German geologist, popularized the term “index fossils” as a marker for the correlation of contemporary strata, based on considerations by Charles Darwin and Charles Lyell concerning the irreversibility of changing morphologies through evolutionary time. Index fossils have to meet a number of requirements: (1) worldwide distribution, (2) rapid development that means a visible short time span of one species, and evident morphological differences in comparison to other species, (3) high number of individuals, (4) high chance of being preserved (Carroll 1988; Ziegler 1992). Obviously, marine organisms (plankton and nekton) are ideally suitable as index fossils, and for that reason paleontologists often focus on aquatic micro- and nanofossils to date sediments (e.g., foraminifera, diatoms). But macrofossils are also appropriate for stratigraphic positioning; these include teeth (Kullmer 1999), and the hard tissues of smaller multicellular species (e.g., ammonites, graptoliths, conodonts). In opposition to index fossils, there are so-called facies fossils that are important for the characterization of a certain milieu with particular ecological conditions; early hominids fit in here. Problems for the chronological determination of strata are produced by geomorphological events like earthquakes, volcanism, dislocations, and disconcordance on the one hand and organism-induced disturbances like bioturbation and digging on the other hand.

A special case in stratigraphy is sequence- or cyclostratigraphy, which corresponds to the global correlation of tectonically independent eustatic sea level fluctuations that happen synchronically in the form of transgressions (sea level raises over supratidal area) and regressions (sea level falls under the lowest tide gage) that are reflected in terms of changing communities (terrestrial/aquatic) and graphical curves. Further methods of dating involve measurements of cyclic events, e.g., seasonal temperature changes, among others: paleomagnetism, palynology (microflora and -fauna, e.g., spores, pollen, ostracods, radiolarians), dendrochronology (counting annual tree rings), dentin annulations (counting tooth dentin layers), diverse luminescence dating methods, stable isotope analysis in ice- or deep-sea cores and microfossils (Lee-Thorp and Sponheimer Volume Chapter 9). Investigations of paleolimnological detritus layers are a method once developed by the Swedish scientist Baron de Geer in 1878 and denominated as the exploration of “varves” (Swedish term that stands for “periodical recurrence”), annual laminated sediments in paleolakes. These methods all miss a link to “absolute” time, i.e., there is, with the exception of tree
## Table 17.5
An overview of dating methods (relative and absolute)

<table>
<thead>
<tr>
<th>Dating methods</th>
<th>Description in brief</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Relative dating methods</strong></td>
<td></td>
</tr>
<tr>
<td>Cation ratio</td>
<td>Geographical dependent age determination of a rock by surface analyses (positive charged ions in the varnish that is formed on a rock), very inaccurate</td>
</tr>
<tr>
<td>Cultural affiliation</td>
<td>Determination of temporal levels of tool-industries, ceramics, etc. made by the temporal community in a certain area</td>
</tr>
<tr>
<td>Fluorine dating</td>
<td>Fluorine accumulates in bone material (fluorapatite) that is deposited in groundwater-leading layers, thus providing information on the past time after burial</td>
</tr>
<tr>
<td>Obsidian hydration analysis (OHA)</td>
<td>Obsidian absorbs atmospheric humidity, old artifacts show a thicker “rind” of hydrate than younger ones (dependent on external factors like: soil type, climate, erosion, burning, etc.)</td>
</tr>
<tr>
<td>Patination</td>
<td>This technique is used when multiple artifacts of the same type are found in the same area and under the same conditions; several kinds of patina are related to time</td>
</tr>
<tr>
<td>Pollen analysis</td>
<td>The study of chronological vegetational history by using microfossils in a target area</td>
</tr>
<tr>
<td>Rate of accumulation</td>
<td>Rock layers accumulate over time, thus, the deeper the layer, the older (applies also for artifacts associated with layers)</td>
</tr>
<tr>
<td>Seriation</td>
<td>Changes in ceramic forms over time to reconstruct consistent patterns of cultural trait-change</td>
</tr>
<tr>
<td>Varve analysis</td>
<td>The thickness and shape of annually laminated sediments in a specific ecosystem (mostly paleolimnological) account for composition, displacement, and climate</td>
</tr>
<tr>
<td>Fossils</td>
<td>“Zones” of organismic remains (mostly species) are determined to classify certain strata</td>
</tr>
<tr>
<td><strong>Absolute dating methods</strong></td>
<td></td>
</tr>
<tr>
<td>Archeomagnetism</td>
<td>Determination of variations (intensity, direction) in the earth’s magnetic field</td>
</tr>
<tr>
<td>Astronomical dating</td>
<td>Analysis of sun’s declination at the solstices</td>
</tr>
<tr>
<td>Dendrochronology</td>
<td>Counting annual tree rings in order to determine the age of wood and reconstruct seasonal conditions</td>
</tr>
<tr>
<td>Electron spin resonance</td>
<td>Artifacts/fossils are exposed to radiation that predictably changes the magnetic field of the object (nondestructable)</td>
</tr>
<tr>
<td>Thermoluminescence dating (TL)</td>
<td>When reheating artifacts, the emitted light of specific crystals is proportional to the amount of radiation absorbed since the material was last heated, thus, provides a method to date pottery, hearths, fire-heated rocks, and burned minerals</td>
</tr>
</tbody>
</table>
Optically stimulated luminescence (OSL)  
Similar to TL, uses light to innervate vacated electrons in sediments, comparisons could be made through sediments with a known amount of added radiation

Fission track  
Uranium (U\(^{238}\) and U\(^{235}\)) radioactive elements create fission tracks by spontaneous splitting of Uranium atoms and therefore exposing high amounts of energy that destroy the crystal lattice in a mineral

Oxidizable Carbon ratio (OCR)  
Soil bodies are analyzed to determine the linear progression of slow humus and charcoal recycling through time with an increase in readily oxidizable carbon and a decrease in the total amount of organic carbon

Potassium–Argon dating (K\(^{40}/\)Ar\(^{40}\))  
The measurement of the accumulation of Argon in a mineral over time

Argon–Argon dating (Ar\(^{39}/\)Ar\(^{40}\))  
Comparison of the amount of Ar\(^{40}\) and Ar\(^{39}\); Ar\(^{40}\) is a stable isotope, thus does not decrease in time, whereas Ar\(^{39}\) is radioactive and consequently decreases over time

Radio-Carbon dating (C\(^{14}\))  
Age estimation for organic materials by measuring the disintegration of radioactive C\(^{14}\) since an organism died, data indication is given by years before present (BP) whereas “present” means exactly AD 1950

Uranium–Thorium dating (U\(^{234}/\)Th\(^{230}\))  
Comparison which uses the properties of radioactive half-life U\(^{234}\) and Th\(^{230}\), measures the equilibrium between these elements and not the accumulation of a decay product

Uranium–Lead dating  
The ratio of radioactive Uranium and Thorium isotopes to Lead as decay product, dating of rock material with a tremendous time depth of billions of years

Thorium–Lead dating  

Rubidium–Strontium dating  
Stable isotope ratio \(^{87}Sr/^{86}Sr\) is measured in rock material

Racemization  
Amino acids as subunits of proteins are widespread in organisms; living organisms contain only L-amino acids (turn polarized light to the left), in dead organisms L-amino acids degrade stepwise to D-amino acids (turn polarized light to the right) until the amounts are equal, thus, the ratio between both conformations provides information about the age of a sample (5,000–ca. 200,000 years depth)

\(^{2}H, ^{13}C, ^{15}N, ^{18}O, ^{34}S\) stable isotopes  
These isotopes are found for several reasons in different amounts in sample materials of different kinds and could be useful for agricultural, archeological, ecological, nutritional, geochemical, or medical research

<table>
<thead>
<tr>
<th>Dating methods</th>
<th>Description in brief</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optically stimulated luminescence (OSL)</td>
<td>Similar to TL, uses light to innervate vacated electrons in sediments, comparisons could be made through sediments with a known amount of added radiation</td>
</tr>
<tr>
<td>Fission track</td>
<td>Uranium (U(^{238}) and U(^{235})) radioactive elements create fission tracks by spontaneous splitting of Uranium atoms and therefore exposing high amounts of energy that destroy the crystal lattice in a mineral</td>
</tr>
<tr>
<td>Oxidizable Carbon ratio (OCR)</td>
<td>Soil bodies are analyzed to determine the linear progression of slow humus and charcoal recycling through time with an increase in readily oxidizable carbon and a decrease in the total amount of organic carbon</td>
</tr>
<tr>
<td>Potassium–Argon dating (K(^{40}/)Ar(^{40}))</td>
<td>The measurement of the accumulation of Argon in a mineral over time</td>
</tr>
<tr>
<td>Argon–Argon dating (Ar(^{39}/)Ar(^{40}))</td>
<td>Comparison of the amount of Ar(^{40}) and Ar(^{39}); Ar(^{40}) is a stable isotope, thus does not decrease in time, whereas Ar(^{39}) is radioactive and consequently decreases over time</td>
</tr>
<tr>
<td>Radio-Carbon dating (C(^{14}))</td>
<td>Age estimation for organic materials by measuring the disintegration of radioactive C(^{14}) since an organism died, data indication is given by years before present (BP) whereas “present” means exactly AD 1950</td>
</tr>
<tr>
<td>Uranium–Thorium dating (U(^{234}/)Th(^{230}))</td>
<td>Comparison which uses the properties of radioactive half-life U(^{234}) and Th(^{230}), measures the equilibrium between these elements and not the accumulation of a decay product</td>
</tr>
<tr>
<td>Uranium–Lead dating</td>
<td>The ratio of radioactive Uranium and Thorium isotopes to Lead as decay product, dating of rock material with a tremendous time depth of billions of years</td>
</tr>
<tr>
<td>Thorium–Lead dating</td>
<td></td>
</tr>
<tr>
<td>Rubidium–Strontium dating</td>
<td>Stable isotope ratio (^{87}Sr/^{86}Sr) is measured in rock material</td>
</tr>
<tr>
<td>Racemization</td>
<td>Amino acids as subunits of proteins are widespread in organisms; living organisms contain only L-amino acids (turn polarized light to the left), in dead organisms L-amino acids degrade stepwise to D-amino acids (turn polarized light to the right) until the amounts are equal, thus, the ratio between both conformations provides information about the age of a sample (5,000–ca. 200,000 years depth)</td>
</tr>
<tr>
<td>(^{2}H, ^{13}C, ^{15}N, ^{18}O, ^{34}S) stable isotopes</td>
<td>These isotopes are found for several reasons in different amounts in sample materials of different kinds and could be useful for agricultural, archeological, ecological, nutritional, geochemical, or medical research</td>
</tr>
</tbody>
</table>
rings, no determination of a “null-varve,” a starting point (Rothe 2000). An overview of relative and absolute dating methods is given in Table 17.5.

17.3.1.5 What context?

Biotic information may be obtained from body fossils that deliver morphological insights (comparative morphology) and from trace fossils that contain ecological information in the form of organic tracks and inorganic marks (ichnology, e.g., footprints of Laetoli). In the case of body fossils, we should pay attention to the immense differences between life, death, and fossil assemblages (Figure 17.10). A biocenosis simply represents all living species, i.e., the extant “community” in an ecosystem; in contrast, paleobiological remains are differentiated in autochthonous (Thanatocenosis = death or indigenous assemblages, in situ communities) and allochthonous (Taphocoenosis = fossil assemblages) extinct communities, whereas the latter could be affected by little transport events.
(parautochthonous) or represent a mixture of local species and species that are brought in. Differences between life and fossil assemblages may result from “displacement” of various kinds (transport, accumulation, disarticulation, etc.), scavenging, destruction of hard materials (bio-erosion, abrasion), obliteration of specific elements (calcite vs. aragonite), etc., all of which may lead to misinterpretation in terms of taxonomy, ecological, and chronological placement of fossils.

However, even in contemporary communities, it is complicated to sort out decisive external and internal factors and to show which weighting each have on specific developments in an ecological system.

17.3.2 Flora, fauna, and a touch on coevolution

There is—and always has been—a constrained relationship between flora and fauna. Since the first algae produced large amounts of oxygen, completely changing the atmosphere, animals and species have had different effects, although one can assume a comprehensive pattern of coevolution. To put it simply, coevolution or mutual selection is the process of reciprocal influence exerted by entities (mostly species) in an ecosystem (Begon et al. 1998), e.g., predator–prey or parasite–host relationships. The association of insects that need nectar and plants requiring pollen transportation presents a common example. In the case of primate evolution, Sussman (1991) hypothesized that primates evolved in conjunction with the radiation of angiosperm plants and developed terminal branch feeding on nectar and flowers (Bloch and Boyer 2002; Silcox et al. Volume 2 Chapter 1). “Sussman has proposed that grasping extremities and nails on the digits evolved for eating fruit on terminal branches of angiosperms” (Sargis 2002 p 1564). Presumably, our own ancestors began to walk upright long before they developed larger brains, since monotonous coverage of the landscape by tropical forest was beginning to fail due to climatic changes from warm and wet to arid and cool conditions in the Late Miocene (Vrba 1985; Bobe et al. 2002; Bobe and Behrensmeyer 2004). Sussman et al. (1985) and Rayner et al. (1993) suggested that, throughout the Pliocene, australopithecines showed arboreal and bipedal tendencies in a mosaic-like habitat that was marked by significant patches of subtropical forest, large areas of grassland, and savannah of all kinds. “Upright posture, large brain, tool making, and other hominin characters as adaptations to a savannah habitat must be rethought, at least in the case of A. africanus” (Rayner et al. 1993 p 228; for further discussions on the origins of bipedality see Harcourt-Smith Volume 3 Chapter 5 and Senut Volume 3 Chapter 6). Other factors, e.g., isolation, niche boundary shifts, and small population size could also be a kick-off for
evolutionary “innovations” as we find them in hominids (Wolpoff 1999; Henke and Rothe 2003). Just a few steps further in our evolutionary history, and in the debate concerning early Homo as well, we are faced with a huge amount of different hypotheses that aim to link environmental constraints to specific adaptations and morphological transformations (Vrba Volume 3 Chapter 4).

Because of the insufficient preservation of plants in toto, caused by increased decomposition in aerobic and particularly tropical conditions, paleobotanists rely heavily on pollen, spore, and phytolith analyses (Jolly et al. 1998; Elenga et al. 2000; African Pollen Database). However, such things can be dispersed well away from their original habitat through wind and water (Andrews 1992). Density and distribution of these palynological units in sediment layers give us an idea about the allocation of specific plant families, spatial scales of retraction or expansion, and potential plant food proposed. For example, isotopic data on hominin diets have shown an involvement with C₄ grass-foods (Lee-Thorp and Sponheimer Volume 1 Chapter 9; Sponheimer and Lee-Thorp Volume 1 Chapter 18). Large herbivores (e.g., gramineous, foliaceous) are bound to the availability, dispersal, and capacity of particular plants; subsequently, hypotheses about migrations of herds and “predator-prey” relationships between the latter, carnivores, and early hominines, may respectively achieve wider temporal and spatial scales. Bovids are the most common faunal element at most Neogene hominid fossil localities and are often used as indicators to understand Plio-Pleistocene hominid ecological and behavioral changes (Vrba 1995; Kappelmann et al. 1997). Besides, the strong interrelations between fossil species, phytogeography, climate, biogeography, and faunal conditions could lead to improved recognitions of “turnover pulses” in ecological networks and reflect the importance of environmental changes on faunal adaptation, selection, and evolution (Foley 1978, 1994, 1995, 1999; Potts 1998a, 1998b; Lahr and Foley 1998; O’Brien and Peters 1999; Owen-Smith 1999; Vrba 1995, 1999, Volume 3 Chapter 4; deMenocal 2004; Hernández Fernández and Vrba 2006).

17.4 Synthesis

Modifying Dobzhansky’s credo (1973) we state that: Nothing in Evolution makes sense except in the light of ecology. Corresponding to this “Ecolution,” we try to depict cross-sections in three time horizons (early–middle–late) of hominid evolution from ca. 7 Ma until today (Figure 17.11). This should provide a basis for showing how many different viewpoints have to be captured when looking at a complex process like hominoid and hominid evolution. It does not
Summary diagram of paleoenvironmental data and hominid evolution, compiled from different authors. There is a clear linkage between hominid evolution, macro- and micromammal report, climatic changes, stable isotope data, and the dispersals within and out of Africa. Seemingly, colder and more arid conditions combined with adaptational changes in hominids initiated several processes that led to the currently accepted patterns of biogeographical processes in hominid evolution. H. erectus was the first hominin species that spreads into the Old World. The taxonomic relationship between late hominin species is by no means completed. The evolution of stone tools comprises the last 2 Myr and leads to an explosion of artistic expression ca. 18 ka.
Paleoecology: an adequate window on the past?
claim to circumscribe all factors in all periods; it can only provide insights and surely leaves many perspectives unseen.

But is there an “ideal” environment that can be hypothesized for the evolution of African hominids? This and other questions are discussed in this section. Two different kinds of hypotheses can be discerned: (1) the habitat-specific and (2) the variability selection hypothesis. The former considers faunal adaptations to a specific environment, while the latter emphasizes the importance of climatic instability as a trigger for adaptive changes.

Modern interpretations support a step-by-step development of drier, cooler, and more open conditions since the Late Miocene. The influence of an arid-adapted fauna on early hominid evolution at the Mid-Pliocene (near 3.2–2.6 Ma) and the aridification shift (Bonnefille 1983; deMenocal 1995; Dupont and Leroy 1995) are viewed especially as catalyst functions for human evolution.

Vrba et al. (1980, 1998, 1995 Volume 3 Chapter 4) is a prominent advocate of the turnover pulse hypothesis that derives from the habitat-specific hypothesis, itself a variation of the savannah hypothesis. Fundamental shifts in African climate—2.8, 1.8, and 1.0 Ma—initiated the so-called “turnovers.” These turnovers are focused bursts of biotic change. For example, between 3.0 and 2.5 Myr many first appearances were of grazing species. This pattern links aridity and expanding grasslands to faunal changes (Vrba 1980; Bobe and Eck 2001). Graphically, the pulses are defined via clustering. Authors, such as Behrensmeyer et al. (1997) or Werdelin and Lewis (2001), do not, however, support this important view of climate- and hominid-evolution interaction.

To make things more complicated, the area for the evolution of bipedality might have been forest habitats, not savannah (Rayner et al. 1993). Concerning the evolution of early tool-making hominids, mosaic zones of grass- and woodland may have stimulated our evolution (Blumenschine 1986).

On the other hand, the variability selection hypothesis advocates the importance of climatic instability for introducing (1) genetic plurality, (2) natural selection, and (3) faunal innovations. Potts (1998b) suggested that many of the largest African faunal evolution events occurred when there were increases in the amplitudes of paleoclimate variability (such as modifications in the durations and amplitudes of orbital-scale wet-dry amplitudes). Potts’ (1998 p 82) view stresses the inconsistency of selection over long time spans and “thus departs from the prevailing paradigm of adaptive evolution via long-term directional selection.” As many biologists take a close match that evolves between an organism and its specific environment (and which can be definitively confirmed as most organisms are indeed habitat specialists), the variability-selection model proposes that lineages experience factors over time that disrupt close connections.
to any specific environment—a decoupling mechanism separating the organism from any environmental state.

17.4.1 Early phase: Forerunners among Miocene primates

In the Late Miocene (23.8–5.3 Ma), the hominoids, including Hylobatidae and Hominidae [Ponginae/Orangutans and Homininae/African Great Apes and Homo; sensu Groves (2001)], had reached their greatest abundance and diversity, and with little doubt human origins lie somewhere within this group (Conroy 1997). Niches now occupied by cercopithecoids were only now additionally exploited by hominoids in Africa (Fleagle 1999). The fossil record of Early, Middle, and Late Miocene is fairly good and includes Proconsulidae and Oreopithecidae as well (Rasmussen Volume 2 Chapter 3; Begun Volume 2 Chapter 4). However, the search for the root of hominid evolution and additionally for “… the identification of hominoids among the various genera and species of fossil apes from that epoch has proved a fruitless exercise thus far” (Fleagle 1999 p 483). Some propose Dryopithecus as a possible predecessor of the clade that includes Great Apes and Humans; others suppose Ouranopithecus to be directly ancestral to later humans. The earliest fossils that are proposed as probable hominid ancestors are Ardipithecus ramidus (ramidus/kadabba) 5.2–4.4 Ma from Ethiopia (White et al. 1994; Haile-Selassie 2001), Orrorin tugenensis ~6 Ma from Kenya (Senut et al. 2001), and Sahelanthropus tchadensis 6–7 Ma from Chad (Brunet et al. 2002). All these specimens were found in the last 10 years probably because excavation campaigns were targeted toward the “roots” of hominids and the divergence point between apes and hominids, respectively. There is much debate about the “ape-or-human” status, the morphological features, and the changes that have to be carried out on the tree of human evolution (Gibbons 2002), and of special interest here are the ecological circumstances around 7–4 Ma that these species confronted.

Ardipithecus is the name given to 5.8–4.4-Myr-old fossils from the Middle Awash area of Ethiopia. Haile-Selassie (2001 p 178) declared Ardipithecus to be the genus that “… postdated the divergence of lineages leading to modern chimpanzees and humans.” The fossils are associated with a relatively wet and wooded paleoenvironment (Haile-Selassie 2001; WoldeGabriel et al. 2001). WoldeGabriel et al. (2001 p 175) suggested that similar habitats were found in the case of Orrorin (see later) and therefore that “… these findings require fundamental reassessment of models that invoke a significant role for global climatic
change and/or savannah habitat in the origin of hominids.” Associated vertebrate fossil assemblages indicate woodland/forest habitats and small areas of open grassland around lake margins. Among the micromammals the rarity of lagomorphs shows that open grasslands are not well sampled in the habitat of *Ardipithecus*. Moreover, closed wooded environments, where fossils are less likely to be preserved, may explain the low numbers of hominoid/hominid fossils in the Late Miocene. Haile-Selassie (2001) assumed that a foot phalanx (AME-VP-1/71) gives a hint of an early form of terrestrial bipedality corresponding to locomotory abilities in *Australopithecus afarensis* and that dental characters like lower canines are exclusively shared with early hominids.

East African *Orrorin* seems to fall somewhere between the African Great Apes and humans and thus “... accords with the East Side Story proposed by Coppens” (Senut et al. 2001 p 142). The “East Side story” is a construct that invokes the geomorphologically induced allopatric development of African Great Apes and hominids (~8 Ma) in the placement of the cradle of mankind in eastern Africa between the Great Rift Valley and Indian Ocean (Coppens 1987, 1999; deMenocal 2004). A general trend of worldwide cooling, the extensive increase of grassland, and the retraction of tropical forests, rainforests, and wooded savannas are perceptible ~8 Ma. At the same time, a rifting process and an uplifting of the western rift shoulder led to the appearance of a topographic borderline that placed them in the more and more arid eastern part in the rainshadow of the wetter western part (Pickford 1990). This eastern “isolation” for several million years might have been behind endemic peripatric genetic drifts and consequently the origin of hominid features (Coppens 1999). After analyzing and interpreting postcraniual morphological features, Pickford et al. (2002) assumed *Orrorin* to have been a habitual biped with the ability to climb trees, and they found several apomorphic characters shared with australopithecines and *Homo* but none with *Pan* or *Gorilla* (Senut Volume 3 Chapter 6). The habitat of *Orrorin* was reconstructed, from faunal remains and geological analyses of the Late Miocene Lukeino Formation in the Tugen Hills of Kenya, as a mosaic of open woodland and forests around a lake. *Orrorin* seems to be a representative of a typical “edge species” that lived on the frontier between environmental units (Sussman and Hart Volume 1 Chapter 23). Associated faunal remains stem from colobines, carnivores, and ungulates (Pickford and Senut 2001).

*S. tchadensis* was found in Central Africa, in the western Djurab desert of Chad (Brunet et al. 2002). The locality is interesting because it lies far west to the Rift. Following the discovery of *A. bahrelghazali* in 1995, there was no further evidence for a western dispersal of hominids (Brunet et al. 1995) until *Sahelanthropus*, whose ecological circumstances are remarkable insofar as they are similar...
to those of the eastern fossil sites. This raises the possibility to analogize evolutionary constraints for hominids in the eastern and western different areas. Vignaud et al. (2002) suggested that *Sahelanthropus* lived close to a vast lake with swamp areas and rivers (inferred from fish and amphibious forms) and not far from a sandy desert. As deduced from basicranial and facial structures, Brunet et al. (2002 p 150) concluded that there are clear similarities between *Sahelanthropus* and “... later fossil homininds that were clearly bipedal.” The faunal record shows animals associated with gallery forest, open grassland, and savannah (primates, rodents, equids, bovids, and carnivores). All these observations led Brunet and colleagues to arrive at the conclusion that *Sahelanthropus* lived in a more mosaic-like habitat than *Ardipithecus*, *Orrorin*, and the australopiths.

Still, all three new genera are represented by little fossil evidence and lack intraspecific comparisons, limiting inferences facilitating placement of these fossils into the ape or the human lineages (Brunet et al. 2002; Wolpoff et al. 2002; Wood 2002; Senut Volume 3 Chapter 6). Although there is a consensus from fossil and molecular evidence that the human lineage diverged from that of the chimpanzees between ~6–8 Ma, many questions linger pertaining both to their taxonomic classification (Begun 2004) and the reconstruction of the paleoenvironments these “ape-men” lived in. “The solution is in the mantra of all paleontologists: We need more fossils!” (Begun 2004 p 1480).

### 17.4.2 Middle phase: “Chewer” and “Thinker”

Around 2.4–2.0 Ma, the genus *Homo* first appears in the fossil record of Africa. Its definition and the establishment of a hypodigm have given rise to a labyrinth of ideas and approaches about the number of species involved (Foley 1991; Tattersall 1992; Collard and Wood Volume 3 Chapter 8), the morphological variability, distinctions from the australopiths, and phylogenetic relationships. The crucial questions regarding the borderline between *Australopithecus* and *Homo*, and exactly which forms gave rise to genus *Homo* and originated the “robust” lineage around 2.5 Ma, are still unanswered. From the ecological point of view the crucial question is: “Why did these two lineages split off?” Vrba (1988) provides an overview of general hypotheses that relate environmental changes to biotic evolution and thus provides a basis to speculate about divergence processes (Table 17.6).

Late Pliocene and Early Pleistocene strata contain a number of hominid fossils that obviously belonged to different morphotypes and ways of living. On the one hand, there are robust forms with large faces, huge supraorbital structures, a small brain, and enormous masticatory capacities; on the other hand are forms with small
faces, larger brains, reduced supraorbital features, and reduced size of upper and lower jaw. The latter are associated with the first 2.4-Ma-old lithic tools (“pebble-tools” and “choppers”) as attributed to the Oldowan industry and “handy man” H. habilis. It remains doubtful that australopithecines maintained the kind of Osteodontokeratic culture (bone, teeth, and horn) that Raymond Dart (1957) proposed.

We focus here on the outstanding trends that divide two temporally sympatric genera and morphological/evolutionary “lineages”: Paranthropus and early Homo. The different ecological niches that could have been occupied by the species of both genera are limited by their basic needs as (1) large mammals, (2) terrestrial primates, (3) dwellers of tropical gallery forests/open savannah- or grassland-habitats, (4) interspecific competitors, and (5) K-strategists (Foley 1978; Henke and Rothe 1999). When reflecting on the possibilities of coping with the environment, and niche-separation, and -expansion, we have to encompass all kinds of “internal” influences like: body size, population dynamics, abilities of locomotion and “thinking,” life history strategies, behavior, and social system. Furthermore, the role of “external” factors such as climate, biogeography, sympatric species, and predator–prey relationships have to be taken into account (Foley 1978).

### Table 17.6

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Refugial vs. biotidal areas</td>
<td>Environmental changes affect two basic kinds of geographic areas differently: the biome resists in a refugium whereas it does not in a biotidal area. Refugia are characterized by the persistence of dominant taxa (new species within these taxa may be added); in contrast, biotidal areas are shaped by the temporary appearance and disappearance of dominant taxa</td>
</tr>
<tr>
<td>Turnover-pulse (local/widespread)</td>
<td>Physical environmental change is required to initiate most speciations, extinctions, and distribution drift. Thus, most lineage turnover has occurred in pulses, near-synchronous across diverse groups of organisms. Changes could either be widespread with independent evidence of environmental change or they are mostly local and form a largely random frequency distribution against time</td>
</tr>
<tr>
<td>Climatic/tectonic initiating cause</td>
<td>A particular environmental cause of turnover was global, or at least widespread, climatic change. Alternatively, the cause could be tectonic and thus turnover signature is appropriately geographically restricted</td>
</tr>
<tr>
<td>Variability selection hypothesis</td>
<td>Advocates the importance of environmental variability for introducing (1) genetic plurality, (2) natural selection, and (3) faunal innovations</td>
</tr>
</tbody>
</table>

Compiled after Vrba 1988; Potts 1998.
About 2.0 Ma, several kinds of hominids settled in north western Kenya and south Ethiopia around Lake Turkana (Tattersall 2002): *Paranthropus boisei*, *P. (Australopithecus) aethiopicus*, *H. (Kenyanthropus) rudolfensis*, *H. (Australopithecus) habilis*, *H. erectus (ergaster)*; the genera and species names in parentheses indicate that the debate about their taxonomic status is still in progress. The Turkana Basin (Figure 17.3), situated in the Great Rift Valley from southern Ethiopia into northern Kenya, covers an area of about 3,600 km$^2$ and represents one of the richest fossiliferous areas in Africa. An exceptional breadth of “mosaic-like” geological and environmental diversity has been investigated by the Koobi Fora Research Project. It reaches from the lacustrine/fluvial sediments of the Omo river channel with gallery forests and swamps, across thorn bush and grassy floodplains strongly influenced by seasonal flooding, to the basin margins with an arid climate and totally different sedimentation regimes (kfpr.com/prehistory_of_koobi_fora). The geological record provides volcanic tephra layers, amenable to chronometric dating, that are associated with unusually well preserved fossils. The faunal record has a nearly gapless temporal as well as a lateral component, making it possible to reconstruct paleohabitats over large areas. The archeological remains include Oldowan, Karari, and Acheulian tools that complement the evidence of the fossil hominins.

The Omo basin sediments, fossils, and pollen provide a wealth of information about the ecological circumstances and the temporal distribution of hominids in this region. *P. boisei* and *P. aethiopicus* are strongly adapted to drier conditions in savannah habitats with gallery-forests or patchy wooden refugia. Craniodental morphology and especially microdental analyses suggest an exclusively low-quality herbivorous subsistence with sometimes coarse gramineous parts, dependent on the availability of food in a highly variable seasonal environment that is characterized by both copiousness and scarcity (Rak 1983; Foley 1987; Kay and Grine 1988; Henke and Rothe 1999). In contrast, gracile australopithecines seem to have preferred more humid habitats with large spots of forest (Coppens 2002) and were able to accommodate to substantial environmental variability and dietary shifts (Teaford and Ungar 2000; Bonnefille et al. 2004).

The most widely studied habitat-specific hypothesis is the savannah hypothesis. Dart (1925) already used the open savannah-model as a tool to explain larger brains and bipedality in early *Homo*. This concept plausible prima facie was not confirmed by the data collected (Leakey and Hay 1979; Cerling 1992; Senut et al. 2001): bipedalism was apparently established million years before the savannah grassland expansion. Early *Homo* lived in fairly open, arid habitats, used an enlarged spectrum of food resources, and is best characterized as an opportunistic and omnivorous forager. Basic sustenance was surely provided by plant food,
but the use of tools made accessible difficult-to-reach high-energy sources like edaphic storage organs (tubers and roots) (Laden and Wrangham 2005), and meat from scavenging complemented their diet (Foley 1987; Blumenschine et al. 1994). This latter is important when looking at coevolutionary processes and sympatric interdependencies between early Homo, herbivores, and carnivores (Turner 1984; Lewis 1997; Brantingham 1998). High-energy-density diets and unstructured feeding patterns (originated by the seasonal availability) still characterize present-day human eating behaviors, although today’s nutrition is largely uncoupled from seasonal cycles (Ulijaszek 2002). Hunting, as Lee and DeVore (1968) suggested, is widely accepted as unlikely in early Homo; it seems that the conception of a klepto-parasite seems to best characterize the real natural situation. Comparative primatological studies of chimpanzee populations and analyses of behavioral ecology and nutrition strategies in current African hunter-gatherer tribes (Lupo 2002; Marlowe 2005) support these assumptions. Vrba (1988 p 422) stated that “... in some respects the Homo lineage evolved toward a greater ecological generalization, while in contrast the ‘robust’ lineage(s) became more specialized on resources prevalent in more open environments.”

In the Middle and Late Pleistocene, the diversity of hominids decreased drastically, the “robust” australopithecines became extinct (Klein 1988) and Homo constituted the only remaining genus in the bottleneck of hominid evolution. The questions of how and why that happened and what processes were involved are manifold (Henke and Rothe 1999; Tattersall 2002; Rightmire Volume 3 Chapter 12).

17.4.3 Late phase: Neanderthals and colonizers

As one of the most discussed topics in human evolution, the Neanderthal enigma is a prominent “problem” of the late phase. d’Errico and Sánchez Goñi (2003) investigated the millennial scale climatic variability of OIS3 in the context of Neanderthal extinction. To the extent that population models seek climate as a triggering factor for the colonization of Europe by anatomically modern humans and the Neanderthal extinction, they appear to be highly contradictory due to (1) the lack of terrestrial continuous and well-dated paleoclimatic sequences, (2) uncertainties in the dating methods, and (3) the doubts about the cultural attribution of archeological layers. These authors therefore reviewed the paleoclimatic OIS3-evidence from Iberia and found a fragmentary, low-resolution, and ill-interpreted record. d’Errico and Sánchez Goñi (2003) concluded that Aurignacian moderns colonized the north of Iberia and France at the onset of the H4 event.
They based their results on a correlation between archeological data from Western Europe and from two IMAGES pollen-rich deep-sea cores. Their scenario favors Neanderthal populations existing in desert-steppe-like environments (made up of *Artemisia*, Chenopodiaceae, and *Ephedra* which characterize the H4 of this area), while the Aurignacian moderns were probably not interested in colonizing these arid Mediterranean biotopes. Anatomically modern humans did this only after the H4 event.

However, Finlayson et al. (2004) pointed out problems with this scenario. They admit that d’Errico and Sánchez Goñi’s (2003) paper is very useful in reinforcing the data showing that climatic events worldwide became increasingly unstable during OIS3, but as inferred by Finlayson et al. (2004 p 1208), there are problems of cause and effect: “We suggest that this is a spurious correlation and that what are being observed are two populations responding to the same variable (or variables) in opposite directions. Since no direct proof of cause and effect between Aurignacians and Neanderthals is advanced, this must remain the most parsimonious explanation.” Finally, they interpret d’Errico and Sánchez Goñi’s (2003) data in the opposite way by concluding that the available information points to climate instability fragmenting Neanderthal populations and emphasizing that not a single piece of evidence exists that demonstrates a competition between Moderns and Neanderthals.

Further questions we have to answer when discussing anatomically modern humans in Eurasia, Australasia, and the New World are: *Why and when did the expansion happen? Who and how many participated and where did they arrive?* Figure 17.12 provides an overview of potential migration waves and the time they probably took place. One consensus seems to crystallize from the fossil record: equipped with a stature shape and size near to ours (Turkana-boy from Nariokotome), large brain, increased mobility, handling hunting weapons (400–ka-old spears of Schöningen/Germany) as well as fire, and being provided with a significant curiosity, *H. erectus* (*H. ergaster*) migrated far from the continent of its birth at around 1.8 Ma (Dmanisi fossils) (Gabunia et al. 2001) and conquered the Old World. Thenceforward there are diverse models/hypotheses that try to explain dispersal patterns over the last 1.5 Myr of human evolution. The recent african origin model (RAO) holds that the biocultural transition from late archaic *H. sapiens* to anatomically modern humans was restricted to Africa, with subsequent dispersion and replacement of *H. erectus* ca. 200 ka. Another hypothesis assumes an anagenetic transition from the early hominines to *H. erectus*, (earlier than 1.5 Ma) followed by dispersal. In the view of Templeton (2002), second and a third expansion (Out-of-Africa “again and again”) shaped the modern human gene pool. The assumption that
H. sapiens developed independently and in parallel in several regions of the world is mainly based on Asian fossils (e.g., Zhoukoudian, upper cave) and is summarized as the “Multiregional continuity model,” or the polyphyletic hypothesis, with gene flow maintaining some genetic homogeneity (Wolpoff 1999, 2002). The “mostly Out-of-Africa hypothesis” (Relethford 1998, 2001) is a combination of the African replacement model with the multiregional model. All these hypotheses involve questions of whether migrating hominines replaced local populations or if they interbred (for further discussions see also Bräuer Volume 3 Chapter 14).

Referring to the questions we formulated at the beginning of this section, we want to add a third one: how did hominines migrate from Africa? In the early phase, hominids seem to have been restricted to certain African habitats (although there were relatively large numbers of “intra-African” dispersals [Strait and Wood 1999]); this restriction lapsed through the Pleistocene and hominids seemed to become more eurytopic and thus became able to tolerate a wider range of conditions (Foley 1978; Vrba 1995, 1999; Lahr and Foley 1998). It is possible to
see an ecological patterning in the colonization of temperate regions even when looking at the range of other species (mostly large mammals) that spread out of Africa alongside the hominids (Turner 1984; Lewis 1997; Brantingham 1998; Strait and Wood 1999). Certain characteristics of large mammals provide the capacity to exploit new environments, including (1) large body size, (2) carnivorous behavior, and (3) sociality in larger groups. Hence, we have to think about the general principles that encourage zoogeographic mobility (Foley 1987; Henke and Rothe 1999):

- **Carnivores** are more eurytopic than herbivores. Meat requires less specialized and locally restricted adaptations than plants, thus carnivores are expected to migrate faster than herbivores, which are restricted to specific plants in a specific area.
- **Exogeny**: the attribute of an organism to forage across a variety of niches, giving it higher tolerance, less specialization, increased interspecific competition.
- **Environmental physiology**: Increasing body size enhances the energetic situation of an organism in terms of the relation between body weight and surface (Bergmann’s rule 1847; Aiello and Wells 2002). Consequently, larger mammals are able to cope more efficiently with temperate conditions. The extremities tend to be reduced in colder environments to diminish frostbite as seen in the distal limb segments of Neanderthals (Allen’s rule 1906). “There is indeed some possibility that the increase in body size associated with *Homo erectus* [. . .] may have at least contributed to the success of human geographical spread” (Foley 1987 p 268).

Additionally, home range size and diet quality seem to be closely related to initial dispersals from Africa (Antón et al. 2002). Although these principles are able to account for the expansion of early hominines, certain problems related to surviving high latitude habitats remain: (1) resource availability is highly seasonal and (2) the annual variation in day length shortens the time for foraging and other activities in winter. Favored strategies here are the extraction of resources yielding high returns and an increase in the efficiency with which resources are extracted. The former is linked to an increase in carnivory and high-energy food, the latter to increasing predatory efficiency and, especially in hominines, improved tool-manufacturing and -using. Additionally, producing complex and efficient stone tools demands advanced cognitive competence and accordingly constitutes an interaction between encephalization and culture (Klein 2000; Wynn 2002; Osvath and Gärdensfors 2005; Biagi Volume 1 Chapter 24; Toth and Schick Volume 3 Chapter 21). Most likely, habitat structure and resource
types were the driving forces that mainly influenced dispersals and the sequence of habitat colonization by hominids (Foley 1987).

17.5 Conclusions

Human evolution is not a “Sonderweg” (exceptional way); it follows strictly the rules of evolutionary biology and constitutes neither a special case nor determines the terminal branch-end in the Tree of Life, let alone creation’s crowning glory (Foley 1978, 2002). It is just the story of balanced interrelationships between environment and large mammalian genera possessing preadaptations that facilitated coping with the conditions during a specific time span. Corresponding to Foley (2002), we assume that the following key events altered the direction of radiations and dispersals in the hominid lineage:

1. Invasion of Africa from Asia of an ancestral lineage of Miocene primates and the outcome of African apes, and “hominids” *Ardipithecus, Orrorin*, and *Sahelanthropus*;
3. Megadonty and masticatory-apparatus increase as adaptation to low-quality nutrition in a mosaic-like environment in the “robust” australopithecines (taxa: *P. (A.) aethiopicus, P. (A.) boisei, P. (A.) robustus*) during the latest Pliocene and Early Pleistocene;
4. Radiation of earliest *Homo* is the most problematic of all because the phylogenetic position is extremely uncertain and variation within the group extensive (taxa: *Kenyanthropus platyops, H. (K.) rudolfensis, H. habilis*). Nevertheless, there is a clear hint toward a trend: brain size increase;
5. The diversification and dispersal of *Homo* 1.75 Ma (taxa: *H. erectus, H. ergaster*) and a second radiation of later forms of *Homo* 0.5 Ma (*H. heidelbergensis*), bound to several changes in dental/cranial morphology and behavior (carnivory, hunting, tradigenesis), the development of technologies (stone tools), and the usage of fire;
6. Enhanced neural capabilities and improvement of lithic and hunting technologies (e.g., projectiles) in the context of interglacial/glacial cycles may have led to repeated dispersals following each other with the outcome of *H. neanderthalensis* and *H. sapiens*;
7. Symbolic thought combined with language and the opening of new resources (e.g., aquatic food) characterizes the radiation and global colonization of *H. sapiens* in the last 100 kyr. Why and how the Neanderthals became extinct is still a matter of debate;

8. Recently, humans seem to have detached themselves from selection-pressures (except regarding their own species and some kinds of pathogens). This, however, brings with it great responsibility to safeguard the global ecosystem and thus to afford further evolutionary development.

We would like to stress that primate and consequently hominid evolution is a complex process that has to be seen from different angles and needs all of the different scientific approaches currently available.

One question remains open: *Is paleoecology an adequate window on the past?* We answer that in the affirmative but, as in any other hypothesis-based scientific field, we realize that without paying attention to possible pitfalls and without interpreting the results in context, we are in danger of simply telling stories. Unfortunately, there are no “laws” in ecological science such as exist in chemistry and physics; there are merely a lot of hypotheses that have to be tested and one concept that is close to a “law”: natural selection (Pianka 1983). Hence, we conclude that paleoecology—when conceived with care—is an important field of evolutionary biological, and especially of paleoanthropological, research. When discussing hominid evolution, and consequently our own evolutionary history, the danger of storytelling is potentially higher because fossils are rare, sample biases high, and not least because hypotheses always stand on shaky ground when the subject is ourselves (White 2000).

Actualism may be a path that leads to “true” inferences when ecological models and model-organisms are selected carefully. For example, *Pan*, as our genetically-closest relative, seems to be a good model from which to derive inferences, but as a restricted occupant of dense and humid forest habitats with significant morphological adaptations it might not be the first choice for reconstructing the selective pressures that affected early humans living “on the edge” between open and closed habitats (see also Susman and Hart Volume 1 Chapter 23).

It is obvious that paleoecology demands a multidisciplinary approach. However, nearly all researchers are specialized in their own “scientific niches” and struggle to keep and defend them. The great challenge ought to be the creation of a “scientific biocenosis” that helps us to understand evolutionary processes in terms of all the phenomena we are able to examine. The “scientific environment,” however, is sometimes harsh, and chilly winds blow sharp waving the banner of publication amount and impact factor. We would hope, although, that a high
adaptational species—such as ourselves—could change and improve this “habi-
tat” to enhance the “ecology” of pure science.

Acknowledgments

Thanks to Ian Tattersall and Winfried Henke, first for inviting us to contribute to their extraordinary compendium, second for encouragement, and last but not least for proofreading; Hartmut Rothe for initiating the conception of this Handbook; Holger Herlyn and Uwe Hoßfeld for useful ideas on literature and citation.

References

Bergmann C (1847) Über die Verhältnisse der Wärmeökonomie der Thiere zu ihrer Grösse. Göttinger Studien 3: 95–108


Foley RA (1991a) Another unique species. Longman scientific & technical, Harlow, New York, p 313


Haeckel EHPA (1866) Generelle Morphologie der Organismen: allgemeine Grundzüge der organischen Fromen-Wissenschaft, mechan-
Hastenrath S (1985) Climate and circulation of the tropics. Reidel D, Boston MA


Schwerdtfeger F (1968) ökologie der Tiere (Ein Lehr- und Handbuch in drei Teilen) – Autökologie. Verlag Paul Parey, Hamburg Berlin, 461 pp


Abstract

Stable isotope ratio analysis is now regularly used to investigate early hominin diets based on the principle that “you are what you eat.” Analysis of collagen from Neanderthals and anatomically modern humans prior to 20 ka has shown them to be significantly enriched in $^{15}$N compared to contemporaneous carnivores and herbivores. This suggests that animal foods were a dominant component of their diets, although it must be borne in mind that collagen $\delta^{15}$N can underemphasize the importance of plant foods. Carbon isotope analysis of the enamel mineral of South African australopiths and early Homo has revealed that these taxa consumed $\sim$30% C$_4$ foods such as tropical grasses, sedges, or animals that ate these foods. Moreover, the australopiths are characterized by remarkably variable $\delta^{13}$C values. Chimpanzees, in contrast, are nearly pure C$_3$ consumers even in environments with abundant C$_4$ vegetation. These data suggest that when confronted with increasingly open areas, chimpanzees continue to exploit the foods that are most abundant in forest environments, whereas australopiths utilized novel C$_4$ resources in addition to forest foods.

18.1 Introduction

The nature of early hominin diets has been the subject of lively debate, and not without good reason (Robinson 1954; Dart 1957; Jolly 1970; Binford 1981; Grine and Kay 1988; Sillen 1992; Lee-Thorp et al. 1994; Stiner 1994; Sponheimer and Lee-Thorp 1999a; Richards et al. 2000; Speth and Tchernov 2001). Most large primates spend at least 50% of their waking hours searching for or consuming food (Altmann and Altmann 1970; Teleki 1981; Goodall 1986). Thus, if we seek to know what “a day in the life” of our ancestors was like, understanding what they ate would represent an enormous step in that direction. Furthermore, diet is considered among the most important factors underlying behavioral and
ecological differences among extant primates (Ungar 1998; Fleagle 1999), and thus the story of our ancestors’ evolving diets is likely to be intertwined with that of how our species, *Homo sapiens*, came to be. We can glean paleodietary information from many sources. Archeological evidence in the form of stone tools and butchered animal bones is one source of dietary information, which tells us about the kinds of animals that hominins utilized as well as their procurement strategies (Binford 1981; Brain 1981; Blumenschine 1987; Stiner 1994; Speth and Tchernov 2001). Yet, such evidence tends to overemphasize the importance of animal foods at the expense of plant foods that make up the bulk of most primate diets (Lee 1979; Eaton and Konner 1985; Milton 2002). Moreover, the first potential hominins (Senut et al. 2001; Brunet et al. 2002; Haile-Selassie et al. 2004) precede the earliest archeological traces by millions of years (Semaw et al. 1997; De Heinzelin et al. 1999), and thus the archeological record remains silent with regard to the diets of the earliest hominins.

As a result, paleoanthropologists have had to look for other sources of paleodietary information. Dental allometry/morphology and microwear have received much attention and provided important insights into the diets of our ancestors (Robinson 1954; Grine 1981, 1986; Grine and Kay 1988; Ungar and Grine 1991; Teaford et al. 2002; Ungar 2004). Still, these techniques have limitations. For instance, the relatively large incisors and bunodont molars of modern *Papio* suggest a frugivorous diet (Hylander 1975; Ungar 1998; Fleagle 1999), and yet many *Papio* populations consume large quantities of grasses, for which they have no apparent dental specializations (Altmann and Altmann 1970; Harding 1976; Dunbar 1983; Strum 1987). Ungar (2004) has also argued that the dental morphology of extant apes and early hominins may tell us more about their fallback foods than it does about their “typical” diets. Dental microwear, in turn, reveals a great deal about the mechanical properties of a primate’s food. Primates that eat hard, brittle foods, such as gray-cheeked mangabeys (*Lophocebus albigena*), have relatively more microscopic pitting on their molars than do those that eat more pliant, tough foods like mountain gorillas (*Gorilla gorilla beringei*) (Grine and Kay 1988; Ungar 1998). However, microwear is quickly obliterated and therefore provides dietary information for a few brief days, which may or may not be representative of an individual’s “average” diet. In addition, soft foods, such as animal flesh, may not always produce recognizable microwear signatures. Thus, even with these important techniques in our paleodietary arsenal, a great deal about the diets of early hominins remains unknown.

Consequently, new paleodietary techniques have emerged in recent years, one of the most important of which is stable isotope analysis. The idea behind this technique is that “you are what you eat.” In other words, the isotopic composition of one’s food is ultimately traceable in one’s tissues. Thus, stable
isotope analysis provides a direct chemical means of investigating the diets of modern and fossil primates. In this chapter, we address the contribution of stable isotope analysis to our understanding of early hominin diets. This chapter is divided roughly into two sections: stable isotope analysis of bone and dentine collagen, which has been used to investigate the diets of our close cousins the Neanderthals and early modern humans (Bocherens et al. 1991, 1999, 2001, 2005; Fizet et al. 1995; Richards et al. 2000, 2001; Pettitt et al. 2003; Schulting et al. 2005), and stable isotope analysis of enamel apatite, which has shed much light on the diets of the South African australopiths and early Homo (Lee-Thorp et al. 1994, 2000; Sponheimer and Lee-Thorp 1999, 2003; van der Merwe et al. 2003; Sponheimer et al. 2005a). We proceed in this reverse chronological order in an effort to trace the general development of paleodietary stable isotope studies: collagen was first utilized in 1977 with enamel apatite studies appearing a decade later. In addition, the temporally and geographically restricted discussion herein (i.e., the emphasis on European Neanderthals and South African australopiths) is more a reflection of the limited degree to which stable isotopes have been used to investigate the diets of Plio-Pleistocene hominins than selective presentation on our part. The only data we have excluded for this chapter are for specimens younger than 20 ka. The stable isotope data for all older hominins, that we are aware of, at least, are discussed herein and can be found in Tables 18.1 and 18.2.

18.2 Using organic material

18.2.1 Collagen

The first paleodietary study using stable isotopes sought to document maize consumption among Native American populations in New York State (Vogel and van der Merwe 1977). This application was made possible by differences in the photosynthetic pathways of plants. Tropical grasses, such as maize, use C₄ photosynthesis, while virtually all other potential plant foods in New York State use C₃ photosynthesis. The C₃-photosynthetic pathway discriminates markedly against ¹³C, and as a consequence, C₃ plants have very depleted ¹³C/¹²C ratios. In contrast, plants that utilize the C₄-photosynthetic pathway discriminate less against ¹³C and are consequently relatively enriched in ¹³C (Smith and Epstein 1971; Figure 18.1). These distinct isotopic signatures are passed down into the tissues of individuals that eat these plants. Thus, individuals who eat large quantities of C₄ plants like maize will be relatively enriched in ¹³C compared to those who eat only C₃ vegetation. Vogel and van der Merwe (1977) analyzed
the bone collagen of individuals ranging in age from \(4,000\) BP to \(500\) BP. They saw little evidence of maize consumption among the oldest individuals but found that maize had become a very important dietary resource by about \(1,000\) BP—sometimes representing up to 50% of an individual’s diet. This study paved the way for a plethora of innovative applications in the following decades which relied on stable isotope abundances in bone collagen; however, since collagen is rarely preserved for more than 10,000 years, these investigations were largely confined to the recent past. Yet, recent studies have shown that, under the right conditions, bone collagen can survive for 200,000 years or more (Jones et al. 2001). Thus, it has proven possible analyze the bone collagen of Late Pleistocene hominins in certain cases.

### 18.2.2 Collagen: methodological considerations

Before proceeding to the hominin data, we will briefly discuss a few relevant methodological considerations. Much has changed since the early days of stable...
Table 18.2
Specimens, taxa, proveniences, δ¹³C, and references for all australopith and *Homo* specimens

<table>
<thead>
<tr>
<th>Specimen</th>
<th>Tooth</th>
<th>Taxon</th>
<th>Provenience</th>
<th>δ¹³C</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>SK1512</td>
<td>P</td>
<td><em>Paranthropus robustus</em></td>
<td>SK1</td>
<td>−8.8</td>
<td>Lee-Thorp et al. (1994)</td>
</tr>
<tr>
<td>SK879</td>
<td>M</td>
<td><em>Paranthropus robustus</em></td>
<td>SK1</td>
<td>−8.5</td>
<td>Lee-Thorp et al. (1994)</td>
</tr>
<tr>
<td>SKX5015</td>
<td>LM₃</td>
<td><em>Paranthropus robustus</em></td>
<td>SK1</td>
<td>−9.6</td>
<td>Lee-Thorp et al. (1994)</td>
</tr>
<tr>
<td>SK878</td>
<td>RP₃</td>
<td><em>Paranthropus robustus</em></td>
<td>SK1</td>
<td>−6.8</td>
<td>Lee-Thorp et al. (1994)</td>
</tr>
<tr>
<td>SK879</td>
<td>M</td>
<td><em>Paranthropus robustus</em></td>
<td>SK1</td>
<td>−8.1</td>
<td>Lee-Thorp et al. (1994)</td>
</tr>
<tr>
<td>SKX1312</td>
<td>LM₁</td>
<td><em>Paranthropus robustus</em></td>
<td>SK2</td>
<td>−8.1</td>
<td>Lee-Thorp et al. (1994)</td>
</tr>
<tr>
<td>SKX333</td>
<td>RM₁</td>
<td><em>Paranthropus robustus</em></td>
<td>SK2</td>
<td>−10.0</td>
<td>Lee-Thorp et al. (1994)</td>
</tr>
<tr>
<td>SKX35025</td>
<td>RM</td>
<td><em>Paranthropus robustus</em></td>
<td>SK3</td>
<td>−7.9</td>
<td>Lee-Thorp et al. (1994)</td>
</tr>
<tr>
<td>SK876</td>
<td>M</td>
<td><em>Paranthropus robustus</em></td>
<td>SK1</td>
<td>−6.7</td>
<td>Lee-Thorp et al. (2000)</td>
</tr>
<tr>
<td>TM 1600</td>
<td>LM₂</td>
<td><em>Paranthropus robustus</em></td>
<td>KB3</td>
<td>−7.9</td>
<td>Sponheimer et al. (2005a)</td>
</tr>
<tr>
<td>SK 19</td>
<td>RM₃</td>
<td><em>Paranthropus robustus</em></td>
<td>SK1</td>
<td>−6.3</td>
<td>Sponheimer et al. (2005a)</td>
</tr>
<tr>
<td>SK 41</td>
<td>LM₃</td>
<td><em>Paranthropus robustus</em></td>
<td>SK1</td>
<td>−6.7</td>
<td>Sponheimer et al. (2005a)</td>
</tr>
<tr>
<td>SK 57</td>
<td>LM₃</td>
<td><em>Paranthropus robustus</em></td>
<td>SK1</td>
<td>−6.5</td>
<td>Sponheimer et al. (2005a)</td>
</tr>
<tr>
<td>SK 14000</td>
<td>LM₃</td>
<td><em>Paranthropus robustus</em></td>
<td>SK1</td>
<td>−5.9</td>
<td>Sponheimer et al. (2005a)</td>
</tr>
<tr>
<td>SK 14132</td>
<td>RM₃</td>
<td><em>Paranthropus robustus</em></td>
<td>SK1</td>
<td>−6.9</td>
<td>Sponheimer et al. (2005a)</td>
</tr>
<tr>
<td>SKW 6</td>
<td>LM₃</td>
<td><em>Paranthropus robustus</em></td>
<td>SK1</td>
<td>−7.0</td>
<td>Sponheimer et al. (2005a)</td>
</tr>
<tr>
<td>SKW 3068</td>
<td>LM₂</td>
<td><em>Paranthropus robustus</em></td>
<td>SK1</td>
<td>−8.1</td>
<td>Sponheimer et al. (2005a)</td>
</tr>
<tr>
<td>SKW 4768</td>
<td>LM₂</td>
<td><em>Paranthropus robustus</em></td>
<td>SK1</td>
<td>−7.4</td>
<td>Sponheimer et al. (2005a)</td>
</tr>
<tr>
<td>MLD 30</td>
<td>RM₁</td>
<td><em>Australopithecus africanus</em></td>
<td>MAK3</td>
<td>−5.6</td>
<td>Sponheimer and Lee-Thorp (1999a)</td>
</tr>
<tr>
<td>MLD 41</td>
<td>M</td>
<td><em>Australopithecus africanus</em></td>
<td>MAK3</td>
<td>−11.3</td>
<td>Sponheimer and Lee-Thorp (1999a)</td>
</tr>
<tr>
<td>MLD12</td>
<td>RM₃</td>
<td><em>Australopithecus africanus</em></td>
<td>MAK3</td>
<td>−7.7</td>
<td>Sponheimer and Lee-Thorp (1999a)</td>
</tr>
<tr>
<td>Site</td>
<td>Member</td>
<td>Age</td>
<td>Species</td>
<td>Member</td>
<td>Age</td>
</tr>
<tr>
<td>--------</td>
<td>--------</td>
<td>-----</td>
<td>-----------------------</td>
<td>--------</td>
<td>-----</td>
</tr>
<tr>
<td>MLD28</td>
<td>RM³</td>
<td></td>
<td>Australopithecus africanus</td>
<td>MAK3</td>
<td>–8.1</td>
</tr>
<tr>
<td>STS 31</td>
<td>RM³</td>
<td></td>
<td>Australopithecus africanus</td>
<td>ST4</td>
<td>–6.8</td>
</tr>
<tr>
<td>STS 32</td>
<td>RM³</td>
<td></td>
<td>Australopithecus africanus</td>
<td>ST4</td>
<td>–7.8</td>
</tr>
<tr>
<td>STS 45</td>
<td>RM²</td>
<td></td>
<td>Australopithecus africanus</td>
<td>ST4</td>
<td>–4.0</td>
</tr>
<tr>
<td>STS 72</td>
<td>RM³</td>
<td></td>
<td>Australopithecus africanus</td>
<td>ST4</td>
<td>–9.7</td>
</tr>
<tr>
<td>STS 2218</td>
<td>M</td>
<td></td>
<td>Australopithecus africanus</td>
<td>ST4</td>
<td>–5.9</td>
</tr>
<tr>
<td>STW 73</td>
<td>RM²</td>
<td></td>
<td>Australopithecus africanus</td>
<td>ST4</td>
<td>–8.8</td>
</tr>
<tr>
<td>STW 276</td>
<td>LM₁</td>
<td></td>
<td>Australopithecus africanus</td>
<td>ST4</td>
<td>–8.0</td>
</tr>
<tr>
<td>STW 252</td>
<td>RM¹</td>
<td></td>
<td>Australopithecus africanus</td>
<td>ST4</td>
<td>–7.4</td>
</tr>
<tr>
<td>STW 211</td>
<td>M</td>
<td></td>
<td>Australopithecus africanus</td>
<td>ST4</td>
<td>–7.3</td>
</tr>
<tr>
<td>STW 304</td>
<td>M</td>
<td></td>
<td>Australopithecus africanus</td>
<td>ST4</td>
<td>–7.4</td>
</tr>
<tr>
<td>STW 14</td>
<td>LM₁</td>
<td></td>
<td>Australopithecus africanus</td>
<td>ST4</td>
<td>–6.7</td>
</tr>
<tr>
<td>STW 315</td>
<td>Ldm₂</td>
<td></td>
<td>Australopithecus africanus</td>
<td>ST4</td>
<td>–5.7</td>
</tr>
<tr>
<td>STW 309b (409)</td>
<td>LM₁</td>
<td></td>
<td>Australopithecus africanus</td>
<td>ST4</td>
<td>–6.1</td>
</tr>
<tr>
<td>STW 229</td>
<td>P</td>
<td></td>
<td>Australopithecus africanus</td>
<td>ST4</td>
<td>–5.8</td>
</tr>
<tr>
<td>STW 303</td>
<td>RM²</td>
<td></td>
<td>Australopithecus africanus</td>
<td>ST4</td>
<td>–4.3</td>
</tr>
<tr>
<td>STW 236</td>
<td>P</td>
<td></td>
<td>Australopithecus africanus?</td>
<td>ST4</td>
<td>–3.7</td>
</tr>
<tr>
<td>STW 213i</td>
<td>LM₁</td>
<td></td>
<td>Australopithecus africanus?</td>
<td>ST4</td>
<td>–1.8</td>
</tr>
<tr>
<td>SK 80/847</td>
<td>P</td>
<td></td>
<td>Early Homo</td>
<td>SK1</td>
<td>–7.1</td>
</tr>
<tr>
<td>SK 27</td>
<td>LM³</td>
<td></td>
<td>Early Homo</td>
<td>SK1</td>
<td>–8.2</td>
</tr>
<tr>
<td>SK 2635</td>
<td>P</td>
<td></td>
<td>Early Homo</td>
<td>SK1</td>
<td>–9.2</td>
</tr>
</tbody>
</table>

In the provenience column, site abbreviations (SK = Swartkrans, MAK = Makapansgat, ST = Sterkfontein, KB = Kromdraai B) are followed by the appropriate Member number.
isotope with. Today, only small samples are required for analysis, and automation has led to significantly increased sample throughput in laboratories around the world. Nevertheless, the basic procedures have remained largely the same, even though the protocols vary somewhat from laboratory to laboratory. Here, we briefly summarize the operational protocols of the laboratories that have analyzed the collagen of early hominins. One must keep in mind, however, that this is meant to be a general summary and not an exhaustive step-by-step explication of analytical procedures.

After surface cleaning bone or dentine samples are demineralized in dilute HCl (0.5–1.0 M) for periods ranging from 20 min to 5 days. Performing this step at low temperature (5°C) is one recent innovation that allows collagen to be extracted from very old, fragile samples (Richards and Hedges 1999; Jones et al. 2001). The insoluble residue may then be soaked in 0.125 N NaOH to remove contaminating humic acids, but as this leads to collagen solubilization and decreased extraction yields, it is no longer favored. Fortunately, neglecting this step appears to have little effect on the resulting stable isotope ratios (Bocherens et al. 1999). The residue is then often gelatinized in weak HCl at 75°C–100°C, filtered, and freeze-dried. A small (~1 mg) sample of this purified collagen is then combusted in an elemental analyzer and the resultant CO₂ and N₂ gases are analyzed for ¹³C/¹²C and ¹⁵N/¹⁴N abundances in a stable isotope ratio mass

---

Figure 18.1

¹³C/¹²C ratios of plants using C₃ (trees, bushes, shrubs, and forbs) and C₄ photosynthesis (grasses and some sedges) in Kruger National Park, South Africa. The boxes represent the 25th–75th percentiles (with the medians as horizontal lines) and the whiskers show the 10th–90th percentiles. The carbon isotope compositions of plants using these different photosynthetic pathways are highly distinct. Maize is a tropical grass that uses C₄ photosynthesis, and thus it has a very different carbon isotope composition than other foods that were consumed by Native Americans in New York State (Vogel and van der Merwe 1977)
spectrometer. Stable light isotope ratios are expressed as δ values in parts per thousand (‰) relative to international standards, which are Pee Dee Belemnite (PDB, a marine carbonate) and atmospheric N₂ for carbon and nitrogen, respectively. Standard deviations of replicate measurements are generally about ±0.1‰ for carbon and ±0.2‰ for nitrogen. Importantly, collagen degradation is known to alter stable isotope ratios significantly. Collagen must generally have a carbon/nitrogen ratio between 2.9 and 3.6, with carbon and nitrogen percentages of at least 15% and 5%, respectively, to produce reliable results (Ambrose 1990).

Our understanding of the relationship between dietary δ¹³C and collagen δ¹³C has improved significantly since the first use of stable isotopes for paleodietary reconstruction. Experimental studies of rodents on controlled diets have shown that collagen δ¹³C is enriched by about +5‰ relative to dietary protein (Ambrose and Norr 1993; Tieszen and Fagre 1993), as dietary amino acids are preferentially utilized for tissue synthesis. In contrast, carbon from dietary carbohydrate and lipid makes much less of a contribution to bone collagen (or indeed hair or muscle, etc.). Consequently, bone collagen δ¹³C (and δ¹⁵N) tells you more about the protein component of an individual’s diet than it does about their “whole” or “bulk” diet. This is significant, as animal foods which are high in protein will be overrepresented in bone collagen at the expense of low-protein vegetable foods. This bias must be borne in mind when interpreting collagen stable isotope data.

18.2.3 The Neanderthals

Bocherens et al. (1991) performed the first stable isotope analysis of a Neanderthal and associated fauna from the site of Marillac in France. This study demonstrated that enough collagen could be extracted from bones more than 40,000 years old, and paved the way for subsequent analyses of Neanderthals from Marillac and Saint-Cesaire (Fizet et al. 1995; Bocherens et al. 2005); Scladina Cave, Awirs Cave, and Betchelal-Roche Cave in Belgium (Bocherens et al. 1999; 2001), and Vindija Cave in Croatia (Richards et al. 2000). Virtually all native European plants use the C₃-photosynthetic pathway. As a result, these plants have similar carbon isotope compositions, although those in densely wooded environments are somewhat depleted in ¹³C due to the canopy effect (Vogel 1978; van der Merwe 1989). Thus, not surprisingly, carbon isotopes revealed little about the diets of Neanderthals, save for the possibility that they utilized few food resources from closed environments (Bocherens et al. 1999; Richards et al. 2000) (Table 18.1).

The nitrogen isotopes in Neanderthal bone collagen, however, proved more revealing. Nitrogen isotopes ratios (¹⁵N/¹⁴N) are known to increase by about 3‰
up every step in the food chain (Minagawa and Wada 1984; Schoeninger and DeNiro 1984; Ambrose and DeNiro 1986). Thus, within a hypothetical food web, if plants have $\delta^{15}N$ values of 0‰, herbivores like reindeer have $\delta^{15}N$ values of about 3‰, while carnivores, such as wolves, have $\delta^{15}N$ values of about 6‰. And although nitrogen isotope distributions in food webs are more complicated than this hypothetical example suggests—due to heterogeneity in plant $\delta^{15}N$ and the disparate physiological adaptations and requirements of mammals (Ambrose and DeNiro 1986; Sealy et al. 1987; Sponheimer et al. 2003a)—the general pattern has been shown to be robust in both terrestrial and marine ecosystems. Thus, analysis of nitrogen isotopes in bone collagen can potentially reveal the trophic level at which a hominin was feeding. This is relevant for investigating Neanderthal diets, as their degree of carnivory and manner of carcass acquisition (hunting or scavenging) have been the subject of considerable debate (Binford 1981; Mellars 1989; Stiner 1994; Speth and Tchernov 2001).

Intriguingly, all published studies have shown that Neanderthals have very high $\delta^{15}N$, well above that of contemporaneous herbivores, such as horses (Equus caballus), reindeer (Rangifer tarandus), and bison (Bison priscus), and similar to that of carnivorous wolves (Canis lupus), hyenas (Crocuta spelaea), and lions (Panthera spelaea) (Bocherens et al. 1991, 1997, 2001, 2005; Fizet et al. 1995; Richards et al. 2000). Indeed, analysis of variance and Fisher’s PLSD test of the combined datasets show that Neanderthal $\delta^{15}N (x = 10.9\%o, S.D. = 1.1, n = 12)$ is not only significantly higher than herbivore $\delta^{15}N (x = 5.8\%o, S.D. = 1.6, n = 132) (p < 0.001)$, but also slightly higher than carnivore $\delta^{15}N (x = 9.7\%o, S.D. = 1.1, n = 26) (p = 0.02)$ (Figure 18.2; Table 18.1). We should note that the combined dataset contains specimens that accumulated over ~70,000 years and in locations throughout Europe, and that such temporal and spatial mixing increases intrataxonomic variability making statistically significant differences between taxa less likely. For instance, Stevens and Hedges (2004) found that European horse (Equus spp.) $\delta^{15}N$ fluctuated by at least 4‰ over the last 40,000 years due to glacially mediated changes in the nitrogen cycle. Hence, it is particularly striking that significant differences are found between Neanderthals, herbivores, and carnivores without controlling for temporal and spatial variation. In general, the $\delta^{15}N$ data appear to suggest that Neanderthals were highly carnivorous and received little of their dietary protein from plant foods (Richards et al. 2000), despite the bias toward animal foods in bone collagen. Enrichment in $^{15}N$ compared to (other) carnivores may suggest a dependence on herbivores with relatively high $\delta^{15}N$, such as mammoths (Mammuthus primigenius), or even the consumption of omnivorous bears (Ursus spp.) (Richards et al. 2000; Bocherens et al. 2001, 2005). While interpretation of these data remains difficult, it is clear that Neanderthal diets were distinct from those of contemporary herbivorous fauna.
This leads inexorably to the question “Were Neanderthal diets distinct from those of their hominin contemporaries?” While it has not yet been possible to compare the stable isotope compositions of Neanderthals and contemporaneous anatomically modern humans (AMH), Richards et al. (2001) were able to analyze nine near-contemporaries from the mid-Upper Paleolithic (~28–20 ka) at Brno-Francouzska and Dolni Vestonice (Czech Republic); Kostenki, Mal’ta, and Sunghir (Russia); and Paviland (Great Britain). They compared these anatomically modern humans to the five Neanderthals that had been published at the time, and argued that the modern humans were even more elevated in $\delta^{15}N$ ($\bar{x} = 12.0\%o$, S.D. = 1.6, $n = 9$). This suggested that anatomically modern humans were also highly dependent on animal foods. Richards et al. (2001) argued, however, that even the consumption of high-$\delta^{15}N$ herbivores would not be sufficient to account for the extremely elevated $\delta^{15}N$ of these AMH. Instead, they suggested that these humans had begun to diversify their resource base to include freshwater aquatic resources, such as fish and waterfowl, which can be more enriched in $^{15}N$ than terrestrial resources (Dufour et al. 1999; Katzenberg and Weber 1999). This suggestion was rather surprising as there is little direct archeological evidence for exploitation of such foods at this time, although such evidence becomes abundant by the late Upper Paleolithic. Yet, with the subsequent publication of seven new Neanderthal specimens (Bocherens et al. 2001, 2005) and two more mid-Upper Paleolithic humans (Pettitt et al. 2003; Schulting et al. 2005), there is no longer any statistically significant difference in the $\delta^{15}N$ of AMH ($\bar{x} = 11.8\%o$, S.D. = 1.6, $n = 11$) and Neanderthals ($p = 0.06$)
t-test; \( p = 0.09 \) Mann–Whitney \( U \)) (Figure 18.3; Table 18.1). There is, however, a small, but significant difference in the \( \delta^{13}C \) of AMH (\( x = -18.7 \% \), S.D. = 0.5, \( n = 11 \)) and Neanderthals (\( x = -20.0 \% \), S.D. = 0.8, \( n = 12 \)) (\( p < 0.01 \) t-test and Mann–Whitney \( U \)), although the meaning of this difference is unclear. A greater reliance on open area resources, such as \(^{13}C\)-enriched reindeer, by AMH is one of many possible explanations. More salient from the perspective of dietary breadth, Levene’s test reveals no significant differences in the isotopic variability of Neanderthals and AMH (\( \delta^{13}C, p = 0.23; \delta^{15}N, p = 0.61 \)), and thus there is little isotopic support for the utilization of novel resources or greater resource breadth in AMH prior to 20 ka.

These stable isotope studies are an important complement to traditional archeological paleodietary studies as they represent a direct measure of the foods that an individual ate, which is independent of the taphonomic biases that bedevil faunal analyses (Lyman 1994). Nevertheless, interpretation of these stable isotope data is not straightforward, and there remain a number of important unanswered questions. For instance, why are both hominins enriched by more than 5\% compared to associated herbivores when an enrichment of about 3\% would be expected for a carnivore? Stated otherwise, why is their \( \delta^{15}N \) significantly higher than that of associated carnivores? As discussed earlier, this may be partially, but not satisfactorily, explained by the consumption of herbivores with...
unusually high $\delta^{15}N$, such as mammoths, or even by the consumption of omnivores and/or aquatic resources. Another possibility, however, is that there is some physiological explanation for their extremely high $\delta^{15}N$. Experimental studies have shown that when herbivores are fed diets with crude protein contents that are much greater than their nutritional requirements, their diet-tissue spacing can become much greater than $3\%$ (Sponheimer et al. 2003a). This implies that if the prevailing environment forced Neanderthals to consume high-protein diets that considerably exceeded their crude protein requirements, their diet-tissue spacing might have exceeded $3\%$, thus artificially increasing their $\delta^{15}N$ compared to other taxa. On the other hand, committed carnivores generally have smaller diet-tissue spacing than herbivores (Robbins et al. 2005). Regardless, the anomalously high $\delta^{15}N$ of mammoths and low $\delta^{15}N$ of cave bears (Bocherens et al. 1997; Ambrose 1998) may also hint at the importance of physiological adaptations in determining an organism’s nitrogen isotope composition.

It is worth noting that while such interpretive difficulties exist, they do not diminish the significance of these studies. Even if the Neanderthals did have artificially increased diet-tissue spacing due to a high-protein intake, it might erase their distinctiveness from other carnivores but would certainly not make them look herbivorous. Thus, there can be little doubt that Neanderthals consumed large quantities of animal foods.

### 18.3 Using inorganic material

#### 18.3.1 Enamel apatite

Older hominin material is not amenable to this form of analysis as bone collagen is rarely preserved from beyond the Late Pleistocene (Jones et al. 2001). However, the carbon isotopes in bone’s mineral component (a biological apatite) can also be used as dietary proxies (Sullivan and Krueger 1981, 1983; Lee-Thorp and van der Merwe 1987). But even though bone mineral clearly persists beyond bone collagen, it can be altered postmortem, often resulting in the loss of the biogenic dietary signal (Lee-Thorp and van der Merwe 1987; Wang and Cerling 1994; Lee-Thorp 2000). This is due to bone’s high organic content, porosity, and small crystal size, which make it susceptible to dissolution/reprecipitation phenomena and facilitates the incorporation of diagenetic carbonate ions (LeGeros 1991; Lee-Thorp 2000; Lee-Thorp and Sponheimer 2003). Thus, bioapatite paleodietary studies were forestalled until Lee-Thorp and van der Merwe (1987) showed that dental enamel from ancient fauna with well-understood diets retained its biogenic isotope signal. For example, they showed that fossil equids, like their
modern counterparts, had C$_4$-dominated signatures, while fossil tragelaphines had C$_3$-dominated signatures like their modern descendants. Since then, numerous empirical and theoretical studies have substantiated this finding (Lee-Thorp et al. 1989a; Wang and Cerling 1994; Sponheimer and Lee-Thorp 1999b; Hoppe et al. 2003; Lee-Thorp and Sponheimer 2003; Trickett et al. 2003). Enamel’s resistance to diagenetic phenomena is conferred by its virtually organic-free and highly crystalline state (LeGeros 1991), which renders it effectively “prefossilized.” Thus, tooth enamel offered the possibility of investigating the diets of Early Pleistocene and even Pliocene hominins.

18.3.2 Enamel bioapatite: methodological considerations

For the reasons outlined above, only tooth enamel is sampled for stable isotope analysis of hominin and non-hominin specimens that are millions of years old. Initially, a sample of 200–400 mg was needed (about half of a baboon’s molar), but advances in mass spectrometry have reduced the necessary sample size to a few milligrams (Lee-Thorp et al. 1997; Sponheimer 1999). As a result, it has become possible to sample teeth while producing little to no readily observable damage. As a result, more teeth have become available for analysis.

We will now give an overview of the recent sampling and pretreatment protocols, which are modified after Lee-Thorp et al. (1997) and Sponheimer (1999). Specimens are given a careful visual inspection prior to sampling, and those that are heavily stained or have mineral inclusions in the potential sampling areas are excluded from analysis. Generally, only permanent dentition is sampled, with a heavy emphasis on late forming teeth such as M$_2$s and M$_3$s. Powdered enamel samples are acquired (usually from a tooth’s buccal surface) using a low-speed rotary drill with a diamond-tipped burr. However, most of the recently sampled hominin teeth had been previously fractured, which allowed sampling between the occlusal surface and the enamel–dentine junction so as to avoid damage to the external surface of teeth. Although only small samples are taken (~2 mg), we generally attempt to sample over as extensive an area as is possible in order to obtain enamel formed over a considerable period of time. The enamel powder is then pretreated with 0.1 M acetic acid for ~10 min to remove diagenetic carbonates and rinsed to neutrality. This is followed with a ~10 min soak in 1.5% sodium hypochlorite to remove organic contaminants, after which the remaining enamel powder is rinsed to neutrality and freeze-dried. It is worth noting that these pretreatment protocols can vary from laboratory to laboratory, and even within laboratories over time. As different pretreatments can lead to
small but significant differences in a sample’s stable isotope composition (espe-
cially oxygen), one must compare stable isotope values for teeth analyzed follow-
ing different protocols with caution. Finally, each sample is placed in an
individual reaction vessel and analyzed for $^{13}$C/$^{12}$C and $^{18}$O/$^{16}$O using an auto-
carbonate device coupled to a stable isotope ratio mass spectrometer. Carbon and
oxygen isotope ratios are expressed as $\delta$ ($^{13}$C, $^{18}$O) values in parts per thousand
(‰) relative to the PDB standard. The standard deviations of replicate measure-
ments are typically 0.1‰ and 0.2‰ for $\delta^{13}$C and $\delta^{18}$O, respectively. Oxygen
isotopes do provide ecological information (Kohn et al. 1996; Sponheimer
and Lee-Thorp 1999c, 2001), although they are much better known as paleocli-
matic proxies (Prentice and Denton 1988; Ayliffe and Chivas 1990). For this
reason, and because of space constraints, we will not discuss them as paleoecolog-
ical indicators here.

The relationship between dietary $\delta^{13}$C and enamel apatite $\delta^{13}$C has been well
studied. Unlike collagen, apatite tends to reflect the $\delta^{13}$C of the “whole” or “bulk”
diet and not just the protein component (Ambrose and Norr 1993; Tieszen and
Fagre 1993). Thus, apatite and bone collagen $\delta^{13}$C provide different perspectives
on an individual’s diet. Indeed, if one wanted to obtain the most complete picture
of an individual’s diet, one should analyze both collagen and apatite, although
this rarely happens in practice. Most important of our purposes here, however, is
that enamel apatite provides a good average dietary signal and will equally reflect
the consumption of vegetable and animal foods. Yet, there, is evidence that the
relationship between dietary $\delta^{13}$C and apatite $\delta^{13}$C is not constant. Rodent
apatite tends to be enriched by about +10‰ compared to diet (Ambrose and
Norr 1993; Tieszen and Fagre 1993), while large mammal apatite is enriched by
about +13‰ (Lee-Thorp et al. 1989b; Cerling and Harris 1999; Passey et al.
2005). These differences must be borne in mind when comparing rodent and
nonrodent $\delta^{13}$C but are of little significance for the present discussion.

### 18.3.3 The australopiths and early Homo

As in the initial bone collagen study by Vogel and van der Merwe (1977), the
australopith studies were based on the distinct isotopic signatures of C$_3$ (trees,
bushes, shrubs, and forbs) and C$_4$ plants (tropical grasses and some sedges).
In the early 1990s, it was generally believed that Australopithecus africanus had a
diet of fleshy fruits and leaves, much like that of the modern chimpanzee (Pan
troglodytes), while Paranthropus robustus consumed more small, hard foods such
as nuts (Grine 1981, 1986; Grine and Kay 1988; Ungar and Grine 1991). As these
are all C$_3$ foods, one would then expect that Australopithecus and Paranthropus
would have $\delta^{13}C$ values indistinguishable from those of C$_3$ browsers and frugivores. Several studies have shown, however, that the $\delta^{13}C$ of both australopiths is very distinct from that of their C$_3$-consuming coevals (Lee-Thorp et al. 1994, 2000; Sponheimer and Lee-Thorp 1999a; van der Merwe et al. 2003; Sponheimer et al. 2005a).

**Figure 18.3** shows the combined australopith dataset from Makapansgat, Sterkfontein, Kromdraai, and Swartkrans, which now consists of 37 hominin specimens (Table 18.2). Analysis of variance and Fisher’s PLSD test show that both *Australopithecus* ($x = -7.1‰$, S.D. = 1.8, $n = 19$) and *Paranthropus* ($x = -7.6‰$, S.D. = 1.1, $n = 18$) are strongly different from contemporaneous C$_3$ ($x = -11.5‰$, S.D. = 1.3, $n = 61$) and C$_4$ consumers ($x = -0.6‰$, S.D. = 1.8, $n = 60$) ($p < 0.0001$), but cannot be distinguished from each other ($p = 0.18$). This distinction between the hominins and other fauna cannot be ascribed to diagenesis, as there is no evidence that browser or grazer $\delta^{13}C$ has been significantly altered, and diagenesis should affect hominins and nonhominin fauna alike. If we take the mean $\delta^{13}C$ of C$_4$- and C$_3$-consuming herbivores as indicative of pure C$_4$ and C$_3$ diets, respectively, this would indicate diets of about 35–40% C$_4$ vegetation for both *Australopithecus* and *Paranthropus*. Thus, both were eating considerable quantities of C$_4$ resources, possibly grasses, sedges, or animals that ate these plants. None of these possibilities were expected, as extant apes are not known to consume these foods to a significant degree (Goodall 1986; McGrew et al. 1982). Indeed, even in environments where C$_4$ foods are readily available, chimpanzee $\delta^{13}C$ does not indicate any C$_4$ consumption (Schoeninger et al. 1999; Carter 2001). This suggests a fundamental niche difference between the australopiths and extant apes, which is not so surprising given the vast differences in their craniodental morphology (Grine 1981; Kay 1985; Teaford et al. 2002; Ungar 2004).

It is worth noting that the evidence of extensive C$_4$ consumption was not the only surprise in the australopith dataset. Indeed, hominin $\delta^{13}C$ turned out to be more variable than virtually all modern and fossil taxa that have been analyzed in South Africa (Lee-Thorp et al. 1994, 2000; Sponheimer et al. 1999a, 2001, 2003b; Codron 2003; van der Merwe et al. 2003). There is considerable evidence that South African australopith habitats became more open between ~3.0 and ~1.7 Ma (Vrba 1980, 1985; Reed 1997; Luyt and Lee-Thorp 2003), and it might be argued that this environmental change forced the australopiths to modify their diets over time, leading to their unusually variable $\delta^{13}C$. Yet, linear regression demonstrates that there is no relationship between hominin $\delta^{13}C$ and time ($p = 0.63$, $R^2 = 0.01$; Figure 18.4), and there are no significant differences in hominin $\delta^{13}C$ between 3.0-Ma Makapansgat Member 3, 2.5-Ma Sterkfontein Member 4, or 1.8-Ma Swartkrans Member 1 (ANOVA, $p = 0.14$). Indeed, what
is most striking about these data is the lack of change in hominin δ\(^{13}\)C in the face of pronounced environmental change. Somewhat paradoxically, however, within any given time period (Member), hominin δ\(^{13}\)C is highly variable. This might simply indicate that the australopiths had wide habitat tolerances and always inhabited a similar range of microhabitats. This would be consistent with Wood and Strait’s (2004) recent suggestion that early hominins were eurytopic rather than ecological specialists.

In the case of *A. africanus*, the variability is so great that one might ask if there are not two ecologically distinct taxa presently commingled within its hypodigm. If one includes the numbers for three teeth (STW 236, STW 213i, STW 207) that are possibly, but not definitively, attributed to *A. africanus* (van der Merwe et al. 2003), then this taxon would range from nearly pure C\(_3\) to nearly pure C\(_4\) diets. Stated otherwise, the range of δ\(^{13}\)C within *A. africanus* (−1.8‰ to −11.3‰) would be nearly as great as the range for ecologically disparate *Papio* and *Theropithecus* combined (+0.4‰ to −12.6‰) (Lee-Thorp et al. 1989). Stable isotopes in and of themselves cannot address the question of *A. africanus* unity, but numerous researchers have suggested that *A. africanus* might demonstrate more morphological variability than would be expected for a single taxon (Kimbel and White 1988; Clarke 1994; Lockwood 1997; Moggi-Cecchi et al. 1998). Hence the possibility of two taxa, one subsisting largely on C\(_4\) foods and the other on C\(_3\) foods, cannot be dismissed.

So how do the australopiths compare with early *Homo*? Lee-Thorp et al. (2000) tested for isotopic differences between *Paranthropus* and rarer early *Homo*
specimens from Member 1 at Swartkrans. They assumed that if *Homo* consumed more animal foods, as was widely held, then its δ^{13}C values should be enriched compared to the australopiths because many savanna animals eat C_4 grasses. Surprisingly, though, *Homo* δ^{13}C was very similar to that of the australopiths and the results must be interpreted in a similar way: roughly 25% early *Homo*’s diet came from C_4 sources that included C_4 grasses, C_4 sedges, C_4-animal products, or some combination of these foods (Lee-Thorp et al. 2000). However, only three *Homo* specimens from one site have been analyzed and published, and thus comparisons with the more numerous australopith data must be viewed with caution. In fact, van der Merwe et al.’s (1999) unpublished carbon isotope data show a strong difference between East African *Paranthropus* and *Homo*, which serves to underscore the limitations of the currently published dataset. For this reason, the ensuing discussion will focus exclusively on the australopith data.

### 18.3.4 Which C_4 foods?

Which C_4 foods did the australopiths utilize? This question is quite significant, as use of these different resources might have a variety of physiological, social, and behavioral implications. For instance, if australopiths had a grass-based diet similar to that of the modern gelada baboon (*Theropithecus gelada*), this would almost certainly indicate that their diets were less nutrient-dense than those of modern apes, possibly placing important limitations on burgeoning hominin brains and sociality (Aiello and Wheeler 1995; Milton 1999). The converse, that australopiths ate diets rich in animal foods, would indicate a leap in dietary quality over modern apes which could have been a crucial step toward hominin encephalization, the development of stone tool industries, and increased social complexity (Milton 1999). Similarly, it has been suggested that consuming the underground storage organs of plants like C_4 sedges would represent an increase in dietary quality over that of extant great apes because they are lower in dietary fiber than ape fallback foods (Conklin-Brittain et al. 2002). We will now discuss the evidence for the consumption of C_4 grasses, C_4 sedges, and animal foods in turn.

#### 18.3.4.1 The case for grasses

Some researchers noted that the robust craniodental anatomy of the australopiths might have been an adaptation for eating grass seeds and roots as do modern gelada baboons (*T. gelada*) (Jolly 1970; Wolpoff 1973). A dental microwear study
of modern geladas showed that their molar microwear is dominated by scratches with little evidence of pitting (Teaford 1992), however, which is quite unlike the heavily pitted molars of the australopiths (Grine 1986; Grine and Kay 1988). This result is hardly surprising, though, as it would seem unlikely that relatively large-brained hominins could be sustained on gelada-like diet (high in fiber, low in protein, and long chain polyunsaturated fatty acids) without supplementation with higher quality foods. Furthermore, the stable isotope results do not indicate a pure C₄ diet, but rather one in which C₄ foods are very important and not exclusive. Therefore, even if the C₄ component did originate from grasses, one would not expect australopiths to have Theropithecus-like microwear.

One might expect, however, that australopiths and savanna baboon populations that consume large quantities of grass seasonally would show similarities in dental microwear (Altmann and Altmann 1970; Harding 1976; Dunbar 1983; Strum 1987), and indeed two recent studies of Papio molar microwear noted a more australopith-like frequency of pitting than was found in Theropithecus (Daegling and Grine 1999; Nystrom et al. 2004). In addition, a recent elemental analysis of australopithecus tooth enamel showed that while Australopithecus, and to a lesser extent Paranthropus, had higher Sr/Ca ratios than carnivores, browsers, and papionins, their Sr/Ca was quite similar to grazers. In fact, the unusual combination of high Sr/Ca and low Ba/Ca in Australopithecus has only been found in modern fauna that heavily utilize the underground portions of grasses such as warthogs (Phacochoerus africanus) and African mole rats (Cryptomys hottentotus) (Sponheimer et al. 2005b). These elemental data are still preliminary and certainly cannot be used to state affirmatively that early hominins consumed grasses. Nevertheless, they are entirely consistent with the possibility, and suggest avenues for future research.

18.3.4.2 The case for sedges

Sedges have also received attention as a potential C₄ food for australopiths. Conklin-Brittain et al. (2002) argued that a trend toward desiccation in the Pliocene eroded forests and ultimately forced australopiths into new, more open habitats. Although the degree, manner, and timing of this deterioration are a matter of some debate, the fact that it occurred is not (Vrba 1985; deMenocal 1995; Feibel 1997). Conklin-Brittain et al. (2002) reasoned that this loss of forest habitat forced australopiths into environments that were most similar to their ancestral forest homes, namely wetlands, swamps, and river margins. Sedges are readily available in these environments and have been argued to be among the possible
sources of the C\textsubscript{4} signal in australopiths (Sponheimer and Lee-Thorp 1999a; Conklin-Brittain et al. 2002). Some sedges have underground storage organs that have protein levels equal to those of most chimpanzee foods (9% crude protein) but much lower fiber levels (16% fiber) than foods consumed by chimpanzees (33%) (Conklin-Brittain et al. 2002). Thus, the regular inclusion of sedges in australopith diets might represent an increase in dietary quality over extant great apes. Equally important, the underground portions of sedges would be relatively inaccessible to most mammals, yet readily accessible to hominins with crude-digging implements (Hatley and Kappelman 1980), making sedges a high-quality resource for which there is very little competition. Such foods might have been particularly important during the dry season when other preferred dietary resources were scarce. Moreover, there is evidence of humans and other primates consuming sedges. Modern humans have consumed sedges like *Cyperus esculentus* and *Cyperus papyrus* for thousands of years (Tackholm and Drar 1973; Defelice 2002). Western lowland gorillas (*Gorilla gorilla gorilla*) have also been observed consuming the pith of sedges, although in small quantities (Doran and McNeilage 1998). Finally, *Australopithecus*’ high Sr/Ca is consistent with the consumption of sedge USOs (Sillen et al. 1995; Sponheimer et al. 2005b), although as previously mentioned, this is also consistent with the consumption of grass USOs.

But how likely is it that the observed C\textsubscript{4} signal in early hominins was derived from C\textsubscript{4} sedges? Although 33% of the world’s sedges use the C\textsubscript{4}-photosynthetic pathway (Sage et al. 1999), it is incorrect to assume that all or even most sedges available to australopiths would have utilized the C\textsubscript{4} pathway. Although 65% of Kenyan sedges reportedly use C\textsubscript{4} photosynthesis (Hesla et al. 1982), only 35% do in South Africa (Stock et al. 2004). More to the point, a study of sedges in riverine habitats similar to those inhabited by australopiths found that only 28% use C\textsubscript{4} photosynthesis (Sponheimer et al. 2005a). Thus, unless the South African australopiths deliberately sought out C\textsubscript{4} sedges, or the distribution of sedges was markedly different during the Pliocene, the australopiths would have had to have had a diet of 100% sedges to come close to producing the observed 35–40% C\textsubscript{4} signal. A slightly more probable scenario than this extreme sedge specialization is that the australopiths deliberately sought out C\textsubscript{4} sedges with particularly well-developed rootstocks such as *C. papyrus*, and thus a diet of 35–40% sedges would have been sufficient to explain the australopith carbon isotope values. Yet, this too is unlikely, for although these highly edible sedges are common in extensive perennial wetlands like the Okavango Delta, they are rare in the woodland/bushland habitats that were inhabited by australopiths (Reed 1997; Peters and Vogel 2005).
All told, the available data suggest that even if sedges did constitute an important resource for early hominins, they were likely supplemented with other C4 foods. It is worth noting, however, that some early hominin habitats in East Africa, such as the wetlands of the Eastern Lacustrine Plain at Olduvai Gorge (Hay 1976; Deocampo et al. 2002), might have been better sources of edible C4 sedges. The extremely enriched δ13C of *Paranthropus boisei* might indicate that C4 sedges were locally abundant and heavily utilized (van der Merwe et al. 1999). Puech et al. (1986) have also suggested that the dental microwear of early East African hominins is consistent with the consumption of such foods.

### 18.3.4.3 The case for animal foods

Animal foods can mean many different things including large and small vertebrates, invertebrates, and even birds’ eggs. These foods can also be acquired in a variety of ways including active hunting of large game, passive scavenging, and gathering of insects and eggs. Although chimpanzees are known to hunt a variety of small vertebrates such as red colubus monkeys (*Piliocolobus badius*) and blue duiker (*Cephalophus monticola*), these are pure C3 consumers (Teleki 1981; Goodall 1986). Therefore, intake of these foods could not contribute to the C4 component of australopith diets. More likely sources of the reported C4 signal include small grass-eating taxa such as hyraxes (*Procavia* spp.) and cane rats (*Thryonomys swinderianus*). The young of larger species would also be tempting targets. For instance, the young of antelope like reedbuck (*Redunca arundinum*) lie hidden and largely helpless for the first several months of life, making them easy prey for enterprising hominins.

Arthropods are also potential C4 foods. Baboons are known to eat grass-eating grasshoppers (*Acrididae*) almost exclusively during temporary gluts (Hamilton 1987). Grass-eating termites represent another intriguing possibility, particularly given recent studies suggesting that bone tools from Swartkrans were used to extract termites from mounds (Backwell and d’Errico 2001). Stable isotope studies of termites in African savanna environments have shown that they could have contributed to the australopiths’ 13C enrichment (Boutton et al. 1983; Sponheimer et al. 2005a). While termites range from nearly pure C3 to pure C4 consumers, the vast majority of savanna termites, even in densely wooded riverine microhabitats, consume significant proportions of C4 foods. In fact, termites throughout Krugerp National Park eat 35% C4 vegetation on average (Sponheimer et al. 2005a) (Figure 18.5). Thus, termite consumption by australopiths in woodland savanna and even in riverine forest would be expected to
impart some C₄ carbon to consumers. On the other hand, the fact that so few termites have a pure C₄ signal makes it unlikely that termite consumption alone was the source of the strong C₄ signal of australopiths, because it would require a diet of nearly 100% termites. Alternatively, if the hominins selectively preyed upon grass-specialist harvester termites (*Trinervitermes*, *Hodotermes*) with virtually pure C₄ diets, a diet of about 35–40% termites would be sufficient to produce the observed hominin carbon isotope ratios. This scenario, however, is highly unlikely because these C₄ termites are much less common than those with mixed C₃/C₄ diets in woodlands today; and while harvester termites are more abundant in open grasslands and during acute droughts (Braack and Kryger 2003), there is no reason to believe that australopiths frequented such open environments or that drought conditions were so preponderant. Moreover, while *Trinervitermes* builds highly visible above ground nests (mounds), *Hodotermes* does not, making it much less conspicuous on the landscape (Carruthers 1997; Stuart and Stuart 2000). Thus, it is possible and even likely that termites contributed in some way to the unusual δ¹³C values of australopiths, but other C₄ resources were almost certainly consumed in considerable quantities.

It has been suggested that hominid dental anatomy was not well-suited for the processing of animal foods (Teaford et al. 2002), but this observation only pertains to a limited class of animal foods. A great many animal foods require little, if any, oral processing. Termites, grasshoppers, ants, grubs, eggs, and a variety of other insect delicacies may be consumed whole, and even small vertebrates can be swallowed whole or in a few pieces (Smithers 1983). Brains, marrow,
and other soft tissues can also be consumed without oral processing. In addition, no experiments have been conducted to investigate the actual oral and/or pre-oral processing necessary to consume the muscle tissue of small vertebrates. Thus, it seems unwise to unduly limit the potential foods for australopiths until such studies have been undertaken. Furthermore, consumption of animal foods is common among mammals without seemingly appropriate dentition. One obvious example is the aardwolf (Proteles cristatus), which consumes hundreds of thousands of termites per night with largely nonfunctional, obsolescent dentition (Smithers 1983). In some cases this apparent disjunction between dental morphology and trophic behavior might result from the dentition being adapted for other, more mechanically challenging foods in an animal’s diet. For example, capuchin monkeys (Cebus apella) have large, bunodont dentition with thick enamel adapted for consuming fruits and hard nuts. Nonetheless, up to 50% of capuchin diets can come from animal foods, although the average is closer to 25% (Rosenberger and Kinzey 1972; Rosenberger 1992; Fleagle 1999). Similarly, Ungar (2004) has argued that among hominoids, differences in dental morphology primarily reflect their multifarious fallback foods rather than their preferred foods during times of plenty.

18.3.5 What does it mean?

All told, we still cannot be certain which C₄ resources were utilized by australopiths. Grass roots, grasshoppers, bird’s eggs, lizards, rodents, and young antelope might have been important C₄ resources, particularly during the dry season when little other food was readily available. Succulent plants like euphorbias (Euphorbiaceae) and aloes (Aloaceae) (which are rare in most woodlands but have δ¹³C values that are sometimes indistinguishable from those of C₄ grasses) are also possibilities; for although they are often poisonous to humans (and presumably chimpanzees), they are occasionally utilized by baboons and humans (Codron 2003; Peters and Vogel 2005). Further work on dental microwear and morphology, elemental analysis, and the potential availability and nutritional properties of foods may make it possible to identify these C₄ resources with greater confidence. At present, however, it seems likely that australopiths utilized a wide variety of these foods.

Despite this uncertainty, we should not lose sight of the most significant aspect of these stable isotope data, namely that australopiths increased their dietary breadth compared to extant apes by consuming novel C₄ resources, regardless of what these resources were. Similar evidence for increased dietary breadth is
also evident in their thicker enamel, larger postcanine dentition, and greater mandibular corpus robusticity, all of which point to the consumption of hard objects beyond the capabilities of extant apes (Teaford et al. 2002). Thus, the fundamental trophic difference between australopiths and extant apes might be that when confronted with increasingly open areas, apes continue to exploit the foods that are most abundant in forest environments (McGrew et al. 1982), whereas australopiths utilized novel C4 resources in addition to forest foods. There would have been a number of advantages to such a dietary strategy. It would have allowed australopiths to survive and even thrive in a much greater variety of habitats than do modern great apes, potentially allowing expansion of their range. Similarly, the increased dietary breadth could have buffered australopiths against climatic change and habitat loss. Another implication of increased dietary flexibility might be decreased foraging time and mobility, allowing for increased social interaction and possibly greater social complexity. This flexibility could also have increased dietary quality over that of extant apes by adding low-fiber underground storage organs and protein- and lipid-rich animal foods to australopith menus. This might have been an important step leading to greater encephalization and development of the genus Homo (Aiello and Wheeler 1995; Milton 1999; Conklin-Brittain et al. 2002).

If increased dietary breadth was a fundamental australopith adaptation, what are we to make of the later robust australopiths (Paranthropus robustus and Paranthropus boisei) in which the dental adaptations reached their most specialized form? They are believed to have been specialist hard object feeders that were eventually replaced by our Homo forebears who, for the first time, had regular access to higher quality animal foods (Aiello and Wheeler 1995; Milton 1999). An alternative explanation, however, is that the robust australopithecines were the quintessence of the trend toward dietary diversity because they could access the foods of their progenitors as well as harder foods. One might then argue that they were supplanted by Homo not because they ate different kinds of foods but because Homo was more efficient at procuring these resources due to increased use of extra-oral processing (e.g., stone tools) and greater planning depth. It is believed, for instance, that early Homo increased access to bone marrow from scavenged or hunted carcasses of medium to large mammals by using stone tools (Blumenschine 1987), whereas there is debate as to whether or not robust australopiths utilized such technology (Leakey et al. 1964; Susman 1988; Semaw et al. 1997; de Heinzelin et al. 1999). Furthermore, where australopiths may have eaten antelope lambs only when they stumbled upon them, Homo may have had superior planning depth and followed female antelope back to their young, capturing them only when the mother left once more.
18.4 Conclusions

Stable isotope studies have made an important contribution to our understanding of early hominin diets. In the case of Neanderthals, stable isotopes strongly support the contention that they were highly carnivorous, primary predators. As for the australopiths, stable isotopes suggest that they broadened the ancestral ape resource base to include C_4 foods, which coupled with bipedalism, allowed them to pioneer increasingly open and seasonal environments. Yet, as is the case with all paleodietary techniques, stable isotopes leave a great many important questions unresolved. Particularly important are the equifinality problems that are common in stable isotope studies. In other words, many different diets can lead to the same stable isotope signature. For instance, carbon isotopes often cannot distinguish between diets as different as folivory, frugivory, and carnivory. And although some progress has been made using oxygen isotopes to break such equifinalities (Sponheimer and Lee-Thorp 1999c; Carter 2001), there is little reason to believe that this problem can be entirely circumvented. In the end, stable isotopes are one tool among many, all of which provide a slightly different window into the diets of our ancestors. Thus, where stable isotopes cannot determine the favored prey of Neanderthals, faunal analysis has much to contribute (Speth and Tchernov 2001), and when carbon isotopes remain silent on the topic of australopith fallback foods, dental morphology may jump fruitfully into the fray (Ungar 2004). Thus, stable isotopes will prove most informative when pursued as part of a larger, integrated paleodietary investigation.

That being said, stable isotopes themselves still have a great deal to tell us about the diets of our ancestors. To date, only South African australopiths, Neanderthals, and AMH have been analyzed for stable isotopes to any significant extent, leaving entire continents and vast periods of time virtually unexplored. Future stable isotope studies will surely bridge these gaps, and in doing so will greatly increase our knowledge of early hominin diets. Yet, we should not be satisfied to merely work our way through previously neglected or unattainable material but should seek to push the limits of the technique itself. Improving our knowledge of the stable isotope compositions of modern plants and mammals, investigating how physiology affects diet-tissue spacing, and refining organic extraction techniques so that nitrogen isotopes can be analyzed for a wider variety of taxa, will not only improve our ability to reconstruct paleodiets, but will also enable us to address questions in ways that were previously unimaginable. Hopefully, such actualistic and experimental work will serve to hone this already exciting paleodietary tool.
References


Carter ML (2001) Sensitivity of stable isotopes (\(\text{^{13}C}\), \(\text{^{15}N}\), and \(\text{^{18}O}\)) in bone to dietary specialization and niche separation among sympatric primates in Kibale National Park, Uganda. Ph.D. dissertation, University of Chicago.


Estimation of Basic Life History Data of Fossil Hominoids

Helmut Hemmer

Abstract

Relationships between the life cycle and body mass, brain mass and relative brain size of extant primates can be used to estimate life history parameters of extinct species. Methods to predict these key variables from available cranial and post-cranial materials of fossil hominoids, especially hominids, are compiled and evaluated. The use of different concepts of scaling relative brain size is discussed. Brain mass and constant of cephalization data were used as the source material for the estimation of the age at eruption of the first lower molar, the age at female sexual maturity, the age at first breeding, and the maximum life span. Such data support the interpretation of the Late Miocene *Sahelanthropus tchadensis* as a taxon possibly related to the hominid stem-species near the splitting of chimpanzee and hominid lines, confirm the fundamental nature of the australopithecines as progressive apes, not as humans, and support the view of a close relationship of the early Lower Pleistocene *Homo* paleopopulation of Dmanisi (Georgia) to the Late Pliocene but not to the Lower Pleistocene African *Homo* populations.

19.1 Introduction

Fossil bones are the material of classic osteological studies as a base of systematic and phylogenetic conclusions as well as the understanding of functional correlations of morphological structures. They also bear information that may be used for a two-step computing procedure to estimate fundamental life history, ecological and behavioral data that open a progressive level of insight into ecological relations.

Life history study is understood as an approach in evolutionary biology, identifying ontogenetic variables and then asking what impact those variables have on population size and composition. Relationships between the life cycle and anatomical variables, such as body size and brain size, are of particular
interest because they can be used to estimate life history parameters of extinct species (Hemmer 1974, 2003; Gingerich et al. 1982; DeRousseau 1990; Harvey 1990; Smith and Tompkins 1995), even if there are also life history components independent of brain and body size (Harvey and Read 1988).

The first step of such life history data calculation is the estimation of body mass as a key variable, supported by the estimation of brain mass, from available fossil remains. The second step then takes advantage of the first level results, to look for estimates of parameters of the individual’s life cycle, as the age at first reproduction or the maximum life span. How this subsequently affects higher levels of organization as, e.g., ecological aspects or population growth and evolution is an additional step of life history consideration (DeRousseau 1990; Hemmer 2003; Kappeler et al. 2003). This chapter focuses on the way to evaluate such basic information concealed in fossil bones, to look for hominoid, especially hominid, life history aspects. It does not deal with life history correlates as left by processes of growth, with “osteobiographic” techniques (Boyde 1990; Bromage 1990).

19.2 Life history correlations

Basic primate life history dimensions were found to correlate with body mass (Harvey and Clutton-Brock 1985; Harvey et al. 1987; Lee and Kappeler 2003) (Table 19.1), with absolute brain size (Sacher 1975; Harvey et al. 1987; Smith 1989; Smith et al. 1995; Deaner et al. 2003), and relative brain size.

Table 19.1
Relationships between life history variables (days) and adult female body mass (FBM, g) in primates

<table>
<thead>
<tr>
<th>Life history variable</th>
<th>r</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gestation length (GL)</td>
<td>0.74</td>
<td>[\log GL = 0.13 \log FBM + 1.775]</td>
</tr>
<tr>
<td>Weaning age (WA)</td>
<td>0.91</td>
<td>[\log WA = 0.56 \log FBM + 0.433]</td>
</tr>
<tr>
<td>Age at maturity, female (AMF)</td>
<td>0.89</td>
<td>[\log AMF = 0.51 \log FBM + 1.253]</td>
</tr>
<tr>
<td>Age at maturity, male (AMM)</td>
<td>0.89</td>
<td>[\log AMM = 0.47 \log FBM + 1.471]</td>
</tr>
<tr>
<td>Age at first breeding, female (AFB)</td>
<td>0.92</td>
<td>[\log AFB = 0.44 \log FBM + 1.638]</td>
</tr>
<tr>
<td>Interbirth interval (IBI)</td>
<td>0.86</td>
<td>[\log IBI = 0.37 \log FBM + 1.464]</td>
</tr>
<tr>
<td>Lifespan (LS)</td>
<td>0.78</td>
<td>[\log LS = 0.29 \log FBM + 2.893]</td>
</tr>
</tbody>
</table>

Harvey and Clutton-Brock (1985: Tables 4, 5) or Harvey et al. (1987: Tables 16.3, 16.4); correlation coefficients based not on the species but on the subfamily level.
Brain size is more closely correlated with several life history variables, such as the age at sexual maturity or at first breeding, than is body size (Harvey et al. 1987). Unfortunately, data compilation in landmark publications (Harvey and Clutton-Brock 1985; Harvey et al. 1987) was confounded by conversion of cranial capacity with brain mass using a relationship $1 \text{ cm}^3 = 1 \text{ g}$ (Smith et al. 1995). Other studies were based on either cranial capacity or brain mass (Sacher 1975; Smith et al. 1995) (Table 19.2). The factor related to brain size was considered to be maturation rate as a whole rather than any one of its aspects (Smith 1989). There is an extremely high correlation of the age at eruption of the first lower molar, as a rather stable marker of growth with relatively low variance, and brain size ($r = 0.98$). When the effect of body mass is held constant in multiple regression, the partial correlation of $M_1$ eruption and adult brain size remains $r = 0.90$ (Smith 1989).

The use of relative brain size (constant of cephalization, see later) instead of absolute brain mass or cranial capacity in life history correlations (Hemmer 1974) eliminates the influence of body size and allows for some further improvement of the correlation coefficient with most variables. Allometric formulas that may be used as predictor equations for the estimation of life history data in fossil hominoid primates were calculated for the present contribution based on brain mass (brain and body mass data: Bauchot and Stephan 1969; life history data: Harvey and Clutton-Brock 1985) to avoid the former 1:1 confusion with cranial capacity data. The results for the relationship of body size, brain size, and coefficient of cephalization with the life history corner-stages age at sexual maturity, age at first reproduction, and maximum life span (data taken from the compilation of Harvey and Clutton-Brock 1985: Table 1), as well as age at first lower molar eruption and at completion of the dentition (data taken from Smith (1989), with modifications by Smith et al. (1995)), are listed in Tables 19.3 and 19.4 and illustrated in Figures 19.1 and 19.2).

### Table 19.2

<table>
<thead>
<tr>
<th>Life history variable</th>
<th>$r$</th>
<th>Equation</th>
<th>Source and sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M_1$ eruption age (years) ($M_1\text{EA}$)</td>
<td>0.98</td>
<td>$\ln M_1\text{EA} = 0.582 \ln \text{CrC} - 2.405$</td>
<td>Smith et al. (1995): 14 anthropoid species</td>
</tr>
<tr>
<td>Lifespan (years) (LS)</td>
<td>0.835</td>
<td>$\log LS = 0.379 \log BrM + 0.640$</td>
<td>Sacher (1975): 43 anthropoid species</td>
</tr>
</tbody>
</table>

CC, cranial capacity (cm$^3$); BrM, brain mass (g).
Table 19.3
New life history variable allometries in anthropoid primates

<table>
<thead>
<tr>
<th>Life history variable</th>
<th>r</th>
<th>n</th>
<th>Equation</th>
<th>Reliability for prediction in hominid evolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age at sexual maturity, female (months) (ASMF)</td>
<td>0.815</td>
<td>23</td>
<td>( \log ASMF = 0.341 \log BM + 0.349 ) (LSR)</td>
<td>103%, useless</td>
</tr>
<tr>
<td></td>
<td>0.878</td>
<td>23</td>
<td>( \log ASMF = 0.418 \log BM + 0.059 ) (RMA)</td>
<td>68%, useless</td>
</tr>
<tr>
<td></td>
<td>0.889</td>
<td>23</td>
<td>( \log ASMF = 0.482 \log BrM + 0.697 ) (LSR)</td>
<td>24%, useless</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( \log ASMF = 0.549 \log BrM + 0.568 ) (RMA)</td>
<td>3%, useful</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( \log ASMF = 0.693 \log CC + 0.889 ) (LSR)</td>
<td>2%, useful</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( \log ASMF = 0.776 \log CC + 0.801 ) (RMA)</td>
<td>15%, less useful</td>
</tr>
<tr>
<td>Age at sexual maturity, male (months) (ASMM)</td>
<td>0.797</td>
<td>13</td>
<td>( \log ASMM = 0.287 \log BM + 0.686 ) (LSR)</td>
<td>No observed value for Homo sapiens</td>
</tr>
<tr>
<td></td>
<td>0.857</td>
<td>13</td>
<td>( \log ASMM = 0.360 \log BM + 0.411 ) (RMA)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.878</td>
<td>13</td>
<td>( \log ASMM = 0.440 \log BrM + 0.919 ) (LSR)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( \log ASMM = 0.514 \log BrM + 0.777 ) (RMA)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( \log ASMM = 0.664 \log CC + 1.062 ) (LSR)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( \log ASMM = 0.756 \log CC + 0.965 ) (RMA)</td>
<td></td>
</tr>
<tr>
<td>Age at first breeding, female (months) (AFB)</td>
<td>0.856</td>
<td>27</td>
<td>( \log AFB = 0.315 \log BM + 0.584 ) (LSR)</td>
<td>84%, useless</td>
</tr>
<tr>
<td></td>
<td>0.898</td>
<td>27</td>
<td>( \log AFB = 0.368 \log BM + 0.385 ) (RMA)</td>
<td>62%, useless</td>
</tr>
<tr>
<td></td>
<td>0.902</td>
<td>27</td>
<td>( \log AFB = 0.445 \log BrM + 0.913 ) (LSR)</td>
<td>15%, less useful</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( \log AFB = 0.495 \log BrM + 0.816 ) (RMA)</td>
<td>&lt;1%, very useful</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( \log AFB = 0.636 \log CC + 1.096 ) (LSR)</td>
<td>3%, useful</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( \log AFB = 0.704 \log CC + 1.023 ) (RMA)</td>
<td>16%, less useful</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------</td>
<td>-----</td>
<td>--------</td>
<td>-------------------------------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td><strong>Lifespan (years) (LS)</strong></td>
<td>0.797</td>
<td>30</td>
<td>log LS = 0.227 log BM + 0.511 (LSR)</td>
<td>74%, useless</td>
</tr>
<tr>
<td></td>
<td>0.843</td>
<td>30</td>
<td>log LS = 0.285 log BM + 0.297 (RMA)</td>
<td>50%, useless</td>
</tr>
<tr>
<td></td>
<td>0.851</td>
<td>30</td>
<td>log LS = 0.322 log BrM + 0.747 (LSR)</td>
<td>24%, useless</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>log LS = 0.382 log BrM + 0.634 (RMA)</td>
<td>4%, useful</td>
</tr>
<tr>
<td><strong>Age at eruption of first lower molar (AME)</strong></td>
<td>0.971</td>
<td>11</td>
<td>log AME = 0.492 log BM – 1.773 (LSR)</td>
<td>37%, useless</td>
</tr>
<tr>
<td></td>
<td>0.985</td>
<td>11</td>
<td>log AME = 0.507 log BM – 1.826 (RMA)</td>
<td>31%, useless</td>
</tr>
<tr>
<td></td>
<td>0.981</td>
<td>11</td>
<td>log AME = 0.589 log BrM – 1.091 (LSR)</td>
<td>4%, useful</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>log AME = 0.598 log BrM – 1.108 (RMA)</td>
<td>6%, still useful</td>
</tr>
<tr>
<td><strong>Age at complete dentition (ACD)</strong></td>
<td>0.966</td>
<td>10</td>
<td>log ACD = 0.533 log BM – 1.347 (LSR)</td>
<td>21%, useless</td>
</tr>
<tr>
<td></td>
<td>0.974</td>
<td>10</td>
<td>log ACD = 0.551 log BM – 1.416 (RMA)</td>
<td>16%, less useful</td>
</tr>
<tr>
<td></td>
<td>0.964</td>
<td>10</td>
<td>log ACD = 0.629 log BrM – 0.588 (LSR)</td>
<td>16%, less useful</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>log ACD = 0.648 log BrM – 0.624 (RMA)</td>
<td>20%, less useful</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>log ACD = 0.848 log CC – 0.286 (LSR)</td>
<td>25%, useless</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>log ACD = 0.880 log CC – 0.321 (RMA)</td>
<td>30%, useless</td>
</tr>
</tbody>
</table>

BM, body mass; BrM, brain mass; CC, constant of cephalization; LSR, least-squares regression; RMA, reduced major axis.

Reliability for prediction in hominid evolution defined on the basis of the prediction error (PE, %) of the estimation for *Homo sapiens*; PE = \(\frac{\text{observed} - \text{predicted}}{\text{predicted}}\) \times 100: PE \leq 1%, very useful; PE > 1 \leq 5%, useful; PE > 5 \leq 10%, still useful; PE > 10 \leq 20%, less useful; PE > 20%, useless. Data base: BM and BrM: Bauchot and Stephan (1969), CC: Hemmer (1971), ASMF, ASMM, AFB and LS: Harvey and Clutton-Brock (1985), AME and ACD: Smith (1989), with modifications by Smith et al. (1995).
19.3 Estimating body mass

Body size cannot be measured directly being an abstract concept not a concrete parameter. Size may variously be recorded, e.g., as head-body length or height at the withers in quadruped mammals or as stature in hominids, but for general use body mass undoubtedly is the most reliable measure and the size variable of choice (Gingerich et al. 1982; Jungers 1988; Ruff and Walker 1993; Aiello and Wood 1994; Hemmer 2004). Skeletal predictors may allow to estimate body mass on principle on the base of dentition (Table 19.5), skull (Table 19.6), and postcranial elements (Tables 19.7 and 19.8).

There exists no direct biomechanical relationship between tooth and cranial variables and body mass (Jungers 1988), but such dimensions nevertheless depend on a general factor of size. Dentition measurements may allow us to predict a genetically programmed frame of body size than the real size an individual reached during its ontogeny. Allometric relationships between tooth size and body mass have been used to reach the goal of body mass estimation in

Table 19.4
New life history variable allometries in hominoid primates

<table>
<thead>
<tr>
<th>Life history variable</th>
<th>r</th>
<th>n</th>
<th>Equation</th>
<th>Reliability for prediction in hominid evolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age at first breeding, female (months) (AFB)</td>
<td>0.837</td>
<td>5</td>
<td>log AFB = 0.268 log BrM + 1.450 (LSR)</td>
<td>Useless: no significant correlation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>log AFB = 0.320 log BrM + 1.314 (RMA)</td>
<td>Useless: no significant correlation</td>
</tr>
<tr>
<td>Life span (years) (LS)</td>
<td>0.938</td>
<td>5</td>
<td>log AFN = 0.380 log CC + 1.563 (LSR)</td>
<td>9%, still useful</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>log AFN = 0.405 log CC + 1.524 (RMA)</td>
<td>6%, still useful</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>log AFB = 0.320 log BrM + 1.314 (RMA)</td>
<td></td>
</tr>
</tbody>
</table>

Figure 19.1
Bivariate log–log plots (lines: least-squares regressions) of female age at first breeding (data by Harvey and Clutton-Brock 1985, Table 1) against body mass (r = 0.856), brain mass (r = 0.898) (both data sets by Bauchot and Stephan 1969), and constant of cephalization (r = 0.902) (data by Hemmer 1971, based on the respective sets of body mass and brain mass)
Estimation of basic life history data of fossil hominoids

- **Body mass (g)**
  - $10^2$, $10^3$, $10^4$, $10^5$
  - $2$, $3$, $4$, $5$, $6$, $7$, $2$, $3$, $4$, $5$, $6$, $7$

- **Age at first breeding, female (months)**
  - $100$, $101$, $102$, $2$, $3$, $4$, $5$, $6$, $7$, $8$, $2$, $3$, $4$, $5$, $6$, $7$, $8$

- **Brain mass (g)**
  - $101$, $102$
  - $2$, $3$

- **Constant of cephalization**
  - $101$, $102$
  - $2$, $3$
extinct nongeneralized primates (Gingerich et al. 1982; Martin 1990; see discussion in Jungers 1990). This approach depends on what evolutionary grade of primate species is used to predict body mass (Conroy 1987), and would be a hazardous venture so far hominid evolution is concerned, in view of considerable changes of the relative size of the masticatory apparatus (Wolpoff 1973; Jungers 1988). Introduction of taxon-specific conversion factors (prediction by tooth size related to prediction by cranial or postcranial dimensions) may help to overcome that problem (as used in felid body mass prediction by Hemmer 2001, 2004). It must also always be taken in mind that dentition-based body mass estimates cannot present real individual life weights but merely provide some idea about a statistic mean to be awaited at a given linear predictor measure within a population in question.

Cranial dimensions should depend more closely than tooth dimensions on individual ontogenetic body size modeling. They allow us to predict some “normal mass” of an individual. Cranial dimensions prove in that generalized function in some cases to be nearly as good or even better mass predictors as are some of the best postcranial ones (Aiello and Wood 1994; Kappelman 1996). Nevertheless, cranial dimensions may also considerably mislead on the other hand. Comparing the average body mass estimated at the base of seven cranial predictors to the actual body mass of the respective species, the average prediction error (PE) was found to be ±32.3%, ranging from 0.5% to 79.3% (Martin 1990).

In limb bones, the genetic influences are supplemented by loading-related stimuli to a higher extant. They give evidence of an individual’s muscular strength and development (Lanyon 1990). Just here is a crucial point in hominoid evolution, where quadrupedal gait changed to bipedal gait or vice versa. Just skeletal structures that bear a direct functional relationship to body mass are to await to mislead in body mass estimation when the pattern of movement and the strains involved changed in that way. Predictor equations calculated for femoral diaphyseal diameters or circumferences, as well as for femoral articular dimensions (Jungers 1988, 1990; Kappelman 1996), also for cross-sectional geometrical data, such as cortical area or area moments of inertia (Ruff 1990), lose their utility when developed with quadrupedal hominoids but used for mass estimation in bipedal hominids [for convincing illustration of these issues see Figures 8.5, 8.7, and 8.8 in Ruff (1990) comparing femoral dimension allometric lines for

Figure 19.2
Bivariate log–log plots (lines: least-squares regressions) of age of eruption of the lower first molar (data by Smith 1989, with modifications by Smith et al. 1995) against body mass ($r = 0.979$), brain mass ($r = 0.985$), and constant of cephalization ($r = 0.981$) (data sources as for Figure 19.1)
### Table 19.5
Dental predictors of body mass (g) in primates, selected for $r \geq 0.95$

<table>
<thead>
<tr>
<th>Dimension (mm²)</th>
<th>$r$</th>
<th>Equation</th>
<th>Source and sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper cheek tooth row length × width (CTL × CTW)</td>
<td>0.95</td>
<td>$\log BM = 2.06 \log(CTL \times CTW) - 1.00$</td>
<td>Martin (1990): 36 nonhuman species</td>
</tr>
<tr>
<td>$P_3$ crown area (length × width) ($P_3$CRA)</td>
<td>0.954</td>
<td>$\ln BM = 1.21 \ln P_3$CRA + 1.21</td>
<td>Gingerich et al. (1982): 43 nonhuman species</td>
</tr>
<tr>
<td>$P_4$ crown area (length × width) ($P_4$CRA)</td>
<td>0.955</td>
<td>$\ln BM = 1.44 \ln P_4$CRA + 4.20</td>
<td>Gingerich et al. (1982): 43 nonhuman species</td>
</tr>
<tr>
<td>$M_1$ crown area (length × width) ($M_1$CRA)</td>
<td>0.967</td>
<td>$\ln BM = 1.49 \ln M_1$CRA + 3.55</td>
<td>Gingerich et al. (1982): 43 nonhuman species</td>
</tr>
<tr>
<td>$M_2$ crown area (length × width) ($M_2$CRA)</td>
<td>0.99</td>
<td>$\ln BM = 1.572 \ln M_2$CRA + 3.39</td>
<td>Conroy (1987): 8 ape species</td>
</tr>
<tr>
<td>$M_2$ crown area (length × width) ($M_2$CRA)</td>
<td>0.96</td>
<td>$\ln BM = 1.570 \ln M_2$CRA + 3.38</td>
<td>Conroy (1987): 43 nonhuman anthropoid species</td>
</tr>
<tr>
<td>$M_2$ crown area (length × width) ($M_2$CRA)</td>
<td>0.96</td>
<td>$\ln BM = 1.438 \ln M_2$CRA + 3.55</td>
<td>Conroy (1987): 43 nonhuman anthropoid species, females</td>
</tr>
<tr>
<td>Lower cheek teeth crown area (LCRA)</td>
<td>0.968</td>
<td>$\ln BM = 1.31 \ln M_1$CRA + 3.92</td>
<td>Gingerich et al. (1982): 43 nonhuman species</td>
</tr>
<tr>
<td>$\Sigma$ lower cheek teeth crown area (LCRA)</td>
<td>0.964</td>
<td>$\ln BM = 1.38 \ln LCRA + 1.67$</td>
<td>Gingerich et al. (1982): 43 nonhuman species</td>
</tr>
</tbody>
</table>
Table 19.6
Cranial predictors of body mass (g) in primates, selected for \( r \geq 0.95 \)

<table>
<thead>
<tr>
<th>Dimension (mm)</th>
<th>( r )</th>
<th>Equation</th>
<th>Source and sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skull length (SL)</td>
<td>0.98</td>
<td>( \log \text{BM} = 3.89 \log \text{SL} - 4.09 )</td>
<td>Martin (1990): 36 nonhuman species, MA</td>
</tr>
<tr>
<td>Bizygomatic width (BZ)</td>
<td>0.98</td>
<td>( \log \text{BM} = 3.77 \log \text{BZ} - 3.19 )</td>
<td>Martin (1990): 36 nonhuman species, MA</td>
</tr>
<tr>
<td>Internal zygomatic length (IZ)</td>
<td>0.96</td>
<td>( \log \text{BM} = 3.26 \log \text{IZ} - 0.96 )</td>
<td>Martin (1990): 36 nonhuman species, MA</td>
</tr>
<tr>
<td>Palate length (PAL)</td>
<td>0.96</td>
<td>( \log \text{BM} = 3.68 \log \text{PAL} - 2.08 )</td>
<td>Martin (1990): 36 nonhuman species, MA</td>
</tr>
<tr>
<td>Occipital condyle area (OCCA)</td>
<td>0.98</td>
<td>( \log \text{BM} = 2.16 \log \text{OCCA} )</td>
<td>Aiello and Wood (1994): 23 simian species, both sexes, LSR</td>
</tr>
<tr>
<td>(LOCC ( \times ) BOCC)</td>
<td>0.96</td>
<td>( \log \text{BM} = 1.61 \log \text{OCCA} - 1.00 )</td>
<td>Aiello and Wood (1994): 23 simian species, both sexes, RMA</td>
</tr>
<tr>
<td>Occipital condyle length (LOCC)</td>
<td>0.96</td>
<td>( \log \text{BM} = 3.75 \log \text{LOCC} - 0.10 )</td>
<td>Aiello and Wood (1994): 23 simian species, both sexes, LSR</td>
</tr>
<tr>
<td>Foramen magnum area (FMA)</td>
<td>0.98</td>
<td>( \log \text{BM} = 2.15 \log \text{FMA} - 1.20 )</td>
<td>Aiello and Wood (1994): 23 simian species, both sexes, LSR</td>
</tr>
<tr>
<td>Foramen magnum length (LFM)</td>
<td>0.96</td>
<td>( \log \text{BM} = 3.22 \log \text{LFM} )</td>
<td>Aiello and Wood (1994): 23 simian species, both sexes, RMA</td>
</tr>
<tr>
<td>Foramen magnum breadth (BFM)</td>
<td>0.97</td>
<td>( \log \text{BM} = 3.83 \log \text{BFM} - 0.57 )</td>
<td>Aiello and Wood (1994): 23 simian species, both sexes, LSR</td>
</tr>
<tr>
<td>Orbital breadth (BORB)</td>
<td>0.96</td>
<td>( \log \text{BM} = 4.20 \log \text{BORB} - 1.89 )</td>
<td>Aiello and Wood (1994): 23 simian species, both sexes, LSR</td>
</tr>
<tr>
<td>Orbital height (HORB)</td>
<td>0.95</td>
<td>( \log \text{BM} = 4.19 \log \text{HORB} - 1.78 )</td>
<td>Aiello and Wood (1994): 23 simian species, both sexes, LSR</td>
</tr>
<tr>
<td></td>
<td>0.98</td>
<td>( \log \text{BM} = 4.40 \log \text{HORB} - 2.14 )</td>
<td>Aiello and Wood (1994): 23 simian species, both sexes, RMA</td>
</tr>
<tr>
<td></td>
<td>0.96</td>
<td>( \log \text{BM} = 4.42 \log \text{HORB} - 2.12 )</td>
<td>Aiello and Wood (1994): 6 hominoid species, both sexes, LSR</td>
</tr>
<tr>
<td></td>
<td>0.961</td>
<td>( \log \text{BM} = 4.53 \log \text{HORB} - 2.29 )</td>
<td>Aiello and Wood (1994): 6 hominoid species, both sexes, RMA</td>
</tr>
<tr>
<td></td>
<td>0.957</td>
<td>( \log \text{BM} = 4.718 \log \text{HORB} - 2.560 )</td>
<td>Kappelman (1996): 18 catarrhine sp. + ssp., both sexes, LSR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \log \text{BM} = 4.915 \log \text{HORB} - 2.841 )</td>
<td>Kappelman (1996): 18 catarrhine sp. + ssp., both sexes, RMA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \log \text{BM} = 4.445 \log \text{HORB} - 2.155 )</td>
<td>Kappelman (1996): 10 hominoid sp. + ssp., both sexes, LSR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \log \text{BM} = 4.657 \log \text{HORB} - 2.472 )</td>
<td>Kappelman (1996): 10 hominoid sp. + ssp., both sexes, RMA</td>
</tr>
</tbody>
</table>

19597
<table>
<thead>
<tr>
<th>Dimension (mm)</th>
<th>r</th>
<th>Equation</th>
<th>Source and sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Orbital area (ORBA)</td>
<td>0.96</td>
<td>$\log BM = 2.14 \log ORBA - 1.94$</td>
<td>Aiello and Wood (1994): 23 simian species, both sexes, LSR</td>
</tr>
<tr>
<td></td>
<td>0.98</td>
<td>$\log BM = 2.22 \log ORBA - 2.16$</td>
<td>Aiello and Wood (1994): 23 simian species, both sexes, RMA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\log BM = 2.47 \log ORBA - 2.92$</td>
<td>Aiello and Wood (1994): 6 hominoid species, both sexes, LSR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\log BM = 2.52 \log ORBA - 3.05$</td>
<td>Aiello and Wood (1994): 6 hominoid species, both sexes, RMA</td>
</tr>
<tr>
<td>Orbital area, measured by</td>
<td>0.987</td>
<td>$\log BM = 2.284 \log OA - 2.239$</td>
<td>Kappelman (1996): 18 catarrhine sp. + ssp., both sexes, LSR</td>
</tr>
<tr>
<td>computer digitising (OA)</td>
<td></td>
<td>$\log BM = 2.313 \log OA - 2.321$</td>
<td>Kappelman (1996): 18 catarrhine sp. + ssp., both sexes, RMA</td>
</tr>
<tr>
<td></td>
<td>0.987</td>
<td>$\log BM = 2.258 \log OA - 2.176$</td>
<td>Kappelman (1996): 10 hominoid sp. + ssp., both sexes, LSR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\log BM = 2.287 \log OA - 2.261$</td>
<td>Kappelman (1996): 10 hominoid sp. + ssp., both sexes, RMA</td>
</tr>
<tr>
<td>Biorbital breadth (BIOR)</td>
<td>0.96</td>
<td>$\log BM = 3.85 \log BIOR - 2.81$</td>
<td>Aiello and Wood (1994): 23 simian species, both sexes, LSR</td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>$\log BM = 3.95 \log BIOR - 2.98$</td>
<td>Aiello and Wood (1994): 23 simian species, both sexes, RMA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\log BM = 4.82 \log BIOR - 4.67$</td>
<td>Aiello and Wood (1994): 6 hominoid species, both sexes, LSR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\log BM = 5.10 \log BIOR - 5.20$</td>
<td>Aiello and Wood (1994): 6 hominoid species, both sexes, RMA</td>
</tr>
<tr>
<td>Biporionic breadth (BPOR)</td>
<td>0.97</td>
<td>$\log BM = 3.32 \log BPOR - 2.07$</td>
<td>Aiello and Wood (1994): 23 simian species, both sexes, LSR</td>
</tr>
<tr>
<td></td>
<td>0.98</td>
<td>$\log BM = 3.42 \log BPOR - 2.25$</td>
<td>Aiello and Wood (1994): 23 simian species, both sexes, RMA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\log BM = 3.77 \log BPOR - 2.95$</td>
<td>Aiello and Wood (1994): 6 hominoid species, both sexes, LSR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\log BM = 3.84 \log BPOR - 3.10$</td>
<td>Aiello and Wood (1994): 6 hominoid species, both sexes, RMA</td>
</tr>
</tbody>
</table>

LSR, least-squares regression; RMA, reduced major axis; MA, major axis.
Table 19.7
Selection of postcranial predictors of body mass in hominoid primates, correlation coefficient $\geq 0.95$

<table>
<thead>
<tr>
<th>Dimension (mm)</th>
<th>$r$</th>
<th>Equation</th>
<th>Source and sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>12th thoracic vertebral body: AP × transverse diameter of the superior surface (THV)</td>
<td>0.968</td>
<td>$\log BM = 1.3782 \log THV - 2.3132$</td>
<td>McHenry (1992): 7 hominoid species, both sexes, LSR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\log BM = 1.4244 \log THV - 2.4440$</td>
<td>McHenry (1992): 7 hominoid species, both sexes, RMA</td>
</tr>
<tr>
<td></td>
<td>0.999</td>
<td>$\log BM = 0.6552 \log THV - 0.2443$</td>
<td>McHenry (1992): <em>Homo sapiens</em>, both sexes, 3 populations, LSR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\log BM = 0.6556 \log THV - 0.2456$</td>
<td>McHenry (1992): <em>Homo sapiens</em>, both sexes, 3 populations, RMA</td>
</tr>
<tr>
<td>5th lumbar vertebral body: AP × transverse diameter of the superior surface (LUV)</td>
<td>0.951</td>
<td>$\log BM = 1.3574 \log LUV - 2.4210$</td>
<td>McHenry (1992): 7 hominoid species, both sexes, LSR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\log BM = 1.4277 \log LUV - 2.6288$</td>
<td>McHenry (1992): 7 hominoid species, both sexes, RMA</td>
</tr>
<tr>
<td></td>
<td>0.983</td>
<td>$\log BM = 1.1593 \log LUV - 1.9630$</td>
<td>McHenry (1992): <em>Homo sapiens</em>, both sexes, 3 populations, LSR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\log BM = 1.1797 \log LUV - 2.0281$</td>
<td>McHenry (1992): <em>Homo sapiens</em>, both sexes, 3 populations, RMA</td>
</tr>
<tr>
<td>Sacral body: AP × transverse diameter of the superior aspect (SAC)</td>
<td>0.968</td>
<td>$\log BM = 1.4991 \log SAC - 2.9735$</td>
<td>McHenry (1992): <em>Homo sapiens</em>, both sexes, 3 populations, LSR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\log BM = 1.5492 \log SAC - 3.1290$</td>
<td>McHenry (1992): <em>Homo sapiens</em>, both sexes, 3 populations, RMA</td>
</tr>
<tr>
<td>Humerus length [M1] (HLEN)</td>
<td>0.98</td>
<td>$\log BM = 2.68 \log HLEN - 1.94$</td>
<td>Aiello (1981): 23 anthropoid species, both sexes, LSR</td>
</tr>
<tr>
<td></td>
<td>0.99</td>
<td>$\log BM = 2.70 \log HLEN - 2.02$</td>
<td>Aiello and Wood (1994): 21 simian species (17: both sexes), LSR</td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>$\log BM = 2.75 \log HLEN - 2.15$</td>
<td>Aiello and Wood (1994): 23 simian species, both sexes, RMA</td>
</tr>
<tr>
<td>Humerus head AP diameter (HHAP)</td>
<td>0.985</td>
<td>$\log BM = 2.7018 \log HHAP - 2.6388$</td>
<td>McHenry (1992): 7 hominoid species, both sexes, LSR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\log BM = 2.7431 \log HHAP - 2.7022$</td>
<td>McHenry (1992): 7 hominoid species, both sexes, RMA</td>
</tr>
<tr>
<td>Humerus midshaft circumference [M8] (HMSC)</td>
<td>0.98</td>
<td>$\log BM = 2.73 \log HMSC - 0.27$</td>
<td>Aiello (1981): 23 anthropoid species, both sexes, LSR</td>
</tr>
<tr>
<td></td>
<td>0.96</td>
<td>$\log BM = 2.595 \log HMSC - 0.113$</td>
<td>Hartwig-Scherer (1993): African apes, 19 individuals, LSR</td>
</tr>
<tr>
<td></td>
<td>0.98</td>
<td>$\log BM = 2.36 \log HMSC - 0.16$</td>
<td>Aiello and Wood (1994): 21 simian species (17: both sexes), RMA</td>
</tr>
<tr>
<td>Humerus minimum circumference [M7] (HMIN)</td>
<td>0.98</td>
<td>$\log BM = 2.67 \log HMIN - 0.18$</td>
<td>Aiello and Wood (1994): 23 simian species, both sexes, RMA</td>
</tr>
<tr>
<td>Humerus distal epiphyseal breadth [M4] (HEPI)</td>
<td>0.98</td>
<td>$\log BM = 2.49 \log HEPI + 0.26$</td>
<td>Aiello and Wood (1994): 21 simian species (17: both sexes), LSR</td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>$\log BM = 2.56 \log HEPI + 0.17$</td>
<td>Aiello and Wood (1994): 23 simian species, both sexes, RMA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\log BM = 2.72 \log HEPI - 0.20$</td>
<td>Aiello and Wood (1994): 4 hominoid species, both sexes, LSR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\log BM = 2.88 \log HEPI - 0.48$</td>
<td>Aiello and Wood (1994): 4 hominoid species, both sexes, RMA</td>
</tr>
</tbody>
</table>
### Table 19.7 (continued)

<table>
<thead>
<tr>
<th>Dimension (mm)</th>
<th>r</th>
<th>Equation</th>
<th>Source and sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Humerus distal joint breadth (HDJT)</td>
<td>0.98</td>
<td>log BM = 2.44 log HDJT + 0.70</td>
<td>Aiello and Wood (1994): 21 simian species (17: both sexes), LSR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>log BM = 2.49 log HDJT + 0.63</td>
<td>Aiello and Wood (1994): 23 simian species, both sexes, RMA</td>
</tr>
<tr>
<td>Humerus distal: capitular height × articular width (ELB)</td>
<td>0.966</td>
<td>log BM = 1.4115 log ELB – 2.4855</td>
<td>McHenry (1992): 7 hominoid species, both sexes, LSR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>log BM = 1.4617 log ELB – 2.6280</td>
<td>McHenry (1992): 7 hominoid species, both sexes, RMA</td>
</tr>
<tr>
<td>Radius head mediolateral diameter (RADH)</td>
<td>0.955</td>
<td>log BM = 1.9910 log RADH – 0.8912</td>
<td>McHenry (1992): Homo sapiens, both sexes, 3 populations, LSR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>log BM = 2.0859 log RADH – 1.0132</td>
<td>McHenry (1992): Homo sapiens, both sexes, 3 populations, RMA</td>
</tr>
<tr>
<td>Radius midshaft circumference (RMSC)</td>
<td>0.97</td>
<td>log BM = 2.826 log RMSC + 0.031</td>
<td>Hartwig-Scherer (1993): African apes, 19 individuals, RMA</td>
</tr>
<tr>
<td>Acetabular capacity (cm³) (ACCA)</td>
<td>0.987</td>
<td>BM = 4.162 ACCA – 2.541</td>
<td>Suzman (1980): 7 chimpanzees</td>
</tr>
<tr>
<td>Acetabulum height (ACET)</td>
<td>0.967</td>
<td>ln BM = 2.8025 ln ACET – 6.6459</td>
<td>Jungers (1988): 7 hominoid species, both sexes</td>
</tr>
<tr>
<td>Femur length (FL)</td>
<td>0.987</td>
<td>log BM = 3.498 log FL – 6.750</td>
<td>Ruff (1990): 4 nonhuman anthropoid species, both sexes</td>
</tr>
<tr>
<td>Femur head AP diameter [M19.3] (APFH)</td>
<td>0.98</td>
<td>log BM = 2.45 log APFH + 0.92</td>
<td>Aiello and Wood (1994): 21 simian species (17: both sexes), LSR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>log BM = 2.50 log APFH + 0.86</td>
<td>Aiello and Wood (1994): 23 simian species, both sexes, RMA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BM = 2.239 APFH – 36.5</td>
<td>Ruff et al. (1997): modern Homo sapiens sample</td>
</tr>
<tr>
<td>Femur head vertical diameter [M18] (VDFH)</td>
<td>0.970</td>
<td>log BM = 2.6465 log VDFH – 2.4093</td>
<td>McHenry (1992): 7 hominoid species, both sexes, LSR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>log BM = 2.7284 log VDFH – 2.5310</td>
<td>McHenry (1992): 7 hominoid species, both sexes, RMA</td>
</tr>
<tr>
<td></td>
<td>0.976</td>
<td>log BM = 1.7125 log VDFH – 1.0480</td>
<td>McHenry (1992): Homo sapiens, both sexes, 3 populations, LSR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>log BM = 1.7538 log VDFH – 1.1137</td>
<td>McHenry (1992): Homo sapiens, both sexes, 3 populations, RMA</td>
</tr>
<tr>
<td></td>
<td>0.98</td>
<td>log BM = 2.44 log VDFH + 0.95</td>
<td>Aiello and Wood (1994): 21 simian species (17: both sexes), LSR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>log BM = 2.53 log VDFH + 0.83</td>
<td>Aiello and Wood (1994): 23 simian species, both sexes, RMA</td>
</tr>
<tr>
<td></td>
<td>0.988</td>
<td>log BM = 2.466 log VDFH + 0.913</td>
<td>Kappelman (1996): 14 catarrhine sp. + ssp., mostly both sexes, LSR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>log BM = 2.497 log VDFH + 0.872</td>
<td>Kappelman (1996): 14 catarrhine sp. + ssp., mostly both sexes, RMA</td>
</tr>
<tr>
<td></td>
<td>0.996</td>
<td>log BM = 2.628 log VDFH + 0.718</td>
<td>Kappelman (1996): 13 catarrhine sp. + ssp., mostly both sexes, without Homo, LSR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>log BM = 2.640 log VDFH + 0.702</td>
<td>Kappelman (1996): 13 catarrhine sp. + ssp., mostly both sexes, without Homo, RMA</td>
</tr>
<tr>
<td>Measurement</td>
<td>Regression Equation</td>
<td>Reference(s)</td>
<td></td>
</tr>
<tr>
<td>---------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Femur head diameter (FHD)</td>
<td>( \ln \text{BM} = 2.6142 \ln \text{FHD} - 5.4282 )</td>
<td>Jungers (1988): 7 hominoid species, both sexes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \ln \text{BM} = 2.9047 \ln \text{FHD} - 6.3233 )</td>
<td>Jungers (1988): 6 nonhuman hominoid species, both sexes</td>
<td></td>
</tr>
<tr>
<td>Femur shaft AP \times\ transverse diameter inferior to the lesser trochanter [M9 \times M10] (FS)</td>
<td>( \log \text{BM} = 1.1823 \log \text{FS} - 1.5745 )</td>
<td>McHenry (1992): 7 hominoid species, both sexes, LSR</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \log \text{BM} = 1.2152 \log \text{FS} - 1.6605 )</td>
<td>McHenry (1992): 7 hominoid species, both sexes, RMA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \log \text{BM} = 0.7927 \log \text{FS} - 0.5233 )</td>
<td>McHenry (1992): <em>Homo sapiens</em>, both sexes, 3 populations, LSR</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \log \text{BM} = 0.8107 \log \text{FS} - 0.5733 )</td>
<td>McHenry (1992): <em>Homo sapiens</em>, both sexes, 3 populations, RMA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \log \text{BM} = 1.475 \log \text{FS} + 0.524 )</td>
<td>Hartwig-Scherer (1993): African apes, 19 individuals, RMA</td>
<td></td>
</tr>
<tr>
<td>Femur shaft circumference inferior the lesser trochanter (FSC)</td>
<td>( \log \text{BM} = 2.862 \log \text{FSC} - 0.779 )</td>
<td>Hartwig-Scherer (1993): African apes, 19 individuals, RMA</td>
<td></td>
</tr>
<tr>
<td>Femur midshaft transverse diameter (FMTD)</td>
<td>( \log \text{BM} = 2.55 \log \text{FMTD} + 1.19 )</td>
<td>Aiello (1981): 23 anthropoid species, both sexes, LSR</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \log \text{BM} = 2.492 \log \text{FMTD} - 1.696 )</td>
<td>Ruff (1990): 5 anthropoid species, both sexes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \log \text{BM} = 2.533 \log \text{FMTD} - 1.737 )</td>
<td>Ruff (1990): 4 nonhuman anthropoid species, both sexes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \log \text{BM} = 2.541 \log \text{FMTD} - 1.793 )</td>
<td>Ruff (1990): African apes and human, both sexes</td>
<td></td>
</tr>
<tr>
<td>Femur midshaft circumference [M8] (FMSC)</td>
<td>( \log \text{BM} = 2.64 \log \text{FMSC} - 0.29 )</td>
<td>Aiello and Wood (1994): 21 simian species (17: both sexes), LSR</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \log \text{BM} = 2.71 \log \text{FMSC} - 0.39 )</td>
<td>Aiello and Wood (1994): 23 simian species, both sexes, RMA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \log \text{BM} = 2.809 \log \text{FMSC} - 0.597 )</td>
<td>Hartwig-Scherer (1993): African apes, 19 individuals, RMA</td>
<td></td>
</tr>
<tr>
<td>Femur biepicondylar \times\ distal shaft AP diameter (FDIST)</td>
<td>( \log \text{BM} = 1.0829 \log \text{FDIST} - 1.8467 )</td>
<td>McHenry (1992): 7 hominoid species, both sexes, LSR</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \log \text{BM} = 1.1271 \log \text{FDIST} - 1.9840 )</td>
<td>McHenry (1992): 7 hominoid species, both sexes, RMA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \log \text{BM} = 0.9600 \log \text{FDIST} - 1.5678 )</td>
<td>McHenry (1992): <em>Homo sapiens</em>, both sexes, 3 populations, LSR</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \log \text{BM} = 0.9921 \log \text{FDIST} - 1.6762 )</td>
<td>McHenry (1992): <em>Homo sapiens</em>, both sexes, 3 populations, RMA</td>
<td></td>
</tr>
<tr>
<td>Femur epicondylar breadth [M21] (FEPIML)</td>
<td>( \log \text{BM} = 2.48 \log \text{FEPIML} + 0.29 )</td>
<td>Aiello and Wood (1994): 21 simian species (17: both sexes), LSR</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \log \text{BM} = 2.53 \log \text{FEPIML} + 0.21 )</td>
<td>Aiello and Wood (1994): 23 simian species, both sexes, RMA</td>
<td></td>
</tr>
<tr>
<td>Femur epicondylar depth [M24] (FEPIAP)</td>
<td>( \log \text{BM} = 2.59 \log \text{FEPIAP} + 0.28 )</td>
<td>Aiello and Wood (1994): 21 simian species (17: both sexes), LSR</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \log \text{BM} = 2.66 \log \text{FEPIAP} + 0.18 )</td>
<td>Aiello and Wood (1994): 23 simian species, both sexes, RMA</td>
<td></td>
</tr>
<tr>
<td>Femur medial condyle posterior width (MCW)</td>
<td>( \ln \text{BM} = 2.1224 \ln \text{MCW} - 2.6824 )</td>
<td>Jungers (1988): 7 hominoid species, both sexes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \ln \text{BM} = 2.1743 \ln \text{MCW} - 2.8023 )</td>
<td>Jungers (1988): 6 nonhuman hominoid species, both sexes</td>
<td></td>
</tr>
<tr>
<td>Femur lateral condyle posterior width (LCW)</td>
<td>( \ln \text{BM} = 1.9335 \ln \text{LCW} - 1.7269 )</td>
<td>Jungers (1988): 7 hominoid species, both sexes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \ln \text{BM} = 2.1865 \ln \text{LCW} - 2.3033 )</td>
<td>Jungers (1988): 6 nonhuman hominoid species, both sexes</td>
<td></td>
</tr>
<tr>
<td>Tibia length (TL)</td>
<td>( \log \text{BM} = 4.123 \log \text{TL} - 7.914 )</td>
<td>Ruff (1990): 4 nonhuman anthropoid species, both sexes</td>
<td></td>
</tr>
<tr>
<td>Dimension (mm)</td>
<td>r</td>
<td>Equation</td>
<td>Source and sample</td>
</tr>
<tr>
<td>---------------</td>
<td>-----</td>
<td>----------------------------------------------------------</td>
<td>-------------------------------------------------------</td>
</tr>
<tr>
<td>Tibia proximal AP × transverse diameter (TPR)</td>
<td>0.973</td>
<td>$\log BM = 1.2770 \log TPR - 2.5918$</td>
<td>McHenry (1992): 7 hominoid species, both sexes, LSR</td>
</tr>
<tr>
<td></td>
<td>0.991</td>
<td>$\log BM = 1.3127 \log TPR - 2.7066$</td>
<td>McHenry (1992): 7 hominoid species, both sexes, RMA</td>
</tr>
<tr>
<td>Tibia distal AP × transverse diameter of the talar facet (TDIST)</td>
<td>0.965</td>
<td>$\log BM = 1.1806 \log TDIST - 1.5390$</td>
<td>McHenry (1992): 7 hominoid species, both sexes, LSR</td>
</tr>
<tr>
<td></td>
<td>0.991</td>
<td>$\log BM = 1.2232 \log TDIST - 1.6493$</td>
<td>McHenry (1992): 7 hominoid species, both sexes, RMA</td>
</tr>
<tr>
<td>Tibia midshaft circumference [M10] (TMSC)</td>
<td>0.97</td>
<td>$\log BM = 2.8400 \log TMSC - 0.43$</td>
<td>Aiello and Wood (1994): 20 simian samples, LSR</td>
</tr>
<tr>
<td>Medial tibial plateau AP diameter [M4a] (TMED)</td>
<td>0.96</td>
<td>$\log BM = 2.4800 \log TMED + 0.88$</td>
<td>Aiello and Wood (1994): 21 simian species (17: both sexes), LSR</td>
</tr>
<tr>
<td>Lateral tibial plateau AP diameter (TLAT)</td>
<td>0.95</td>
<td>$\log BM = 2.6300 \log TLAT + 0.79$</td>
<td>Aiello and Wood (1994): 23 simian species, both sexes, RMA</td>
</tr>
<tr>
<td>Tibia proximal breadth [M3] (TPROX)</td>
<td>0.98</td>
<td>$\log BM = 2.4500 \log TPROX + 0.35$</td>
<td>Aiello and Wood (1994): 21 simian species (17: both sexes), LSR</td>
</tr>
<tr>
<td>Distal tibial articulation AP diameter (DTB)</td>
<td>0.957</td>
<td>$\ln BM = 2.5037 \ln DTB - 3.9397$</td>
<td>Jungers (1988): 7 hominoid species, both sexes</td>
</tr>
<tr>
<td></td>
<td>0.988</td>
<td>$\ln BM = 2.8561 \ln DTB - 4.8747$</td>
<td>Jungers (1988): 6 nonhuman hominoid species, both sexes</td>
</tr>
</tbody>
</table>

LSR, least-squares regression; RMA, reduced major axis.

Dimensions in mm (or mm²); numbers in square brackets refer to measurements numbers in Martin (Martin and Saller 1959). Body mass in g; equations in italics, body mass in kg.
Homo sapiens with those for pongids or nonhuman primates in general. Mass estimates based on femoral dimensions of a bipedal hominoid using predictor equations developed with a nonhuman hominoid sample should thus be awaited not to meet the real values (44% and 30% overestimates for human males and females with a catarrhine femoral head diameter regression: Kappelman 1996). Femur-based estimates gained with the use of equations calculated with hominoids including man should be shifted somewhat more in the direction of the true values but still differ from them. In the case of using pure Homo sapiens based predictor equations for body mass estimation in early hominids, caution needs to be exercised too, as all fossil hominids seem to have been characterized by considerable skeletal robusticity with thick long-bone cortices (Jungers 1988; McHenry 1992; Ruff and Walker 1993; Ruff et al. 1997).

If the fossil record of single Homo skeletons is complete enough to know body height (stature) and width (bi-iliac breadth), relatively accurate body mass estimations will be possible based on these dimensions (Ruff and Walker 1993; Ruff et al. 1997) (Table 19.8), provided that there are no basic differences in body proportions.

A crude approach to body mass of fossil hominoids other than simple comparisons among hominoid skeletons (for compilation of earlier trials of that type see Suzman 1980) may be done on the basis of close correlations of some cranial and postcranial dimensions with partial skeletal mass to represent body size (Steudel 1980). This procedure allows to compare fossils with the body mass of large, slightly larger than average, average, slightly smaller than average, or small male or female actual ape species. Approaches to body mass estimation in early hominids by a prediction of body height on the basis of postcranial elements and the subsequent use of a body height to body weight ratio (Wolpoff 1973) must meet the requirements of the same stature class, for which the ratio was calculated, and of the same limb to body proportions. Finally, another

<table>
<thead>
<tr>
<th>Table 19.8</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Body mass prediction based on stature and bi-iliac breadth</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Equation</th>
<th>r</th>
<th>Source and sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>BM = 0.413 ST + 2.892 BI – 84.8</td>
<td>0.941</td>
<td>Ruff and Walker (1993): 21 males, world-wide adult population</td>
</tr>
<tr>
<td>BM = 0.498 ST + 1.877 BI – 74.6</td>
<td>0.965</td>
<td>Ruff and Walker (1993): 306 male Karkar Islanders, 8–67 year</td>
</tr>
<tr>
<td>BM = 0.373 ST + 3.033 BI – 82.5</td>
<td>0.90</td>
<td>Ruff et al. (1997): males</td>
</tr>
</tbody>
</table>

Only samples with $r \geq 0.90$ were selected.

BM, body mass (kg); ST, stature (cm); BI, bi-iliac breadth (cm): before applying these formulas, skeletal bi-iliac breadth is converted to living bi-iliac breadth using the equation living BI = 1.17 skeletal BI – 3 (Ruff et al. 1997).
simplified attempt to estimate body mass of early hominids exclusively at the base of the cube of limb bone length or the square of diameter percentage deviations from a *Homo sapiens* standard skeleton (Krantz 1977) may not result in much more trustful body mass estimates than heavier, much heavier, or lighter, much lighter than the standard, as such a procedure neglects size-dependent allometric shifts and evolutionary changes in body proportions.

It should always be kept in mind that body mass estimates resulting from either cranial or postcranial variables are subject to considerable error (Jungers 1990; Martin 1990; Ruff and Walker 1993; Aiello and Wood 1994; Kappelman 1996). Given this fact (Table 19.5) and the use of equations based on highly correlating samples ($r \geq 0.95$; the correlation coefficient depends on the overall size variability and allows for higher values in all primate samples than in taxon specific samples, but the latter otherwise may have a higher misleading potential for predicting equations used for extinct species), different available line-fitting techniques to create predicting equations (least-squares regression, reduced major axis, major axis) and the partial availability of correction factors to compensate for bias, introduced when a body mass estimate derived in logarithmic units is detransformed back to arithmetic units (Smith 1993), as well as many other issues around estimation of body mass in paleontology (Smith 2002), may not really merit that high attention for practical use as sometimes postulated (Gingerich et al. 1982; McHenry 1992; Hartwig-Scherer 1993; Aiello and Wood 1994; Kappelman 1996). Individual body mass estimates of fossil specimens of course may be calculated that are mathematically quite correct, even to the nearest 100 g, for further statistical use; but nevertheless it is biologically wise to list them not in that accurate form but as roughly rounded values. To give a comparable example, rounding to the nearest kilogram was proposed for body mass prediction in fossil cats only for specimens weighing less than 20 kg, to the nearest 5 kg for the range between 20 and 50 kg, to the nearest 10 kg for specimens weighing more than 50 kg (Hemmer 2004).

### 19.4 Estimating brain mass

Cranial capacities measured in cm$^3$ are not the same as brain masses measured in g. In primates, the general relationship will be cranial capacity $>$ brain mass $>$ brain volume (Smith et al. 1995). Unfortunately, this is not always done in comparative publications [e.g., cranial capacities labeled brain mass in Table 1 of Harvey and Clutton-Brock (1985), identical with Table 15-1 of Harvey et al. (1987); see Smith et al. 1995]. There is also no stable relationship between brain mass and cranial capacity [e.g., 1/1.05 as used for humans by different
authors (Smith et al. 1995) or 1/1.14, as used by Hartwig-Scherer (1993), following Count (1947)] over the whole range of brain size variability. The repeatedly found allometric exponent of 1.02 (Martin 1990, for primates; Röhrs and Ebinger 2001, for mammals other than primates) does not mean an isometric relationship between the two variables at correlation coefficients as high as 0.995–0.997 (Table 19.9a). Therefore, all comparative work using brain size should either center on cranial capacities or brain masses. Calculations of relative brain size (cephalization, encephalization quotient) based on body mass clearly should proceed with brain mass. Unfortunately, the available predictive equations for primates or mammals in general give quite differing brain mass values (Table 19.9b). The primate cranial capacity brain mass conversion formula (Martin 1990) is retained here, since it neither produces clearly excessive values in the upper

**Table 19.9a**

Interspecific allometric relationship of cranial capacity and brain mass in primates and in mammals other than primates

<table>
<thead>
<tr>
<th>Sample</th>
<th>Allometric equation</th>
<th>r</th>
<th>authors</th>
<th>Formula converted for mass estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>33 primate species</td>
<td>CrC = 0.94 BrM(^{1.02})</td>
<td>0.996</td>
<td>Martin (1990, p 363 + fig.8.4)</td>
<td>log BrM = 0.98 log CrC + 0.0246</td>
</tr>
<tr>
<td>27 primate species</td>
<td>-</td>
<td>0.995</td>
<td>Ruff et al. (1997)</td>
<td>log BrM = 0.976 log CrC + 0.0596</td>
</tr>
<tr>
<td>17 mammal species</td>
<td>CrC = 0.96 BrM(^{1.02})</td>
<td>0.997</td>
<td>Röhrs and Ebinger (2001)*</td>
<td>log BrM = 0.98 log CrC + 0.0168</td>
</tr>
</tbody>
</table>

*The authors present this equation as log CrC = –0.0015 + 1.02 log BrV (BrV, brain volume) and add the equation BrV = BrM : 1.036. BrM, brain mass; CrC, cranial capacity (cm\(^3\)).

**Table 19.9b**

Cranial capacity based brain mass estimates (g) by the use of different equations (Table 19.9a)

<table>
<thead>
<tr>
<th>Cranial capacity (cm(^3))</th>
<th>Brain mass estimated by the primate formula of Martin (1990)</th>
<th>Brain mass estimated by the primate formula of Ruff et al. (1997)</th>
<th>Brain mass estimated by the mammalian formula of Röhrs and Ebinger (2001)</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>49</td>
<td>52</td>
<td>48</td>
</tr>
<tr>
<td>100</td>
<td>97</td>
<td>103</td>
<td>95</td>
</tr>
<tr>
<td>500</td>
<td>467</td>
<td>494</td>
<td>459</td>
</tr>
<tr>
<td>1000</td>
<td>922</td>
<td>972</td>
<td>905</td>
</tr>
<tr>
<td>1500</td>
<td>1371</td>
<td>1444</td>
<td>1347</td>
</tr>
</tbody>
</table>
Homo sapiens range in view of such conversion data as presented by earlier authors [compilation in Martin and Saller (1959)], nor nonsense values in the lower range, whereby brain volumes would become just larger than cranial capacities (as produced with the use of the formula of Ruff et al. 1997).

It should be noted that there are some pitfalls in the determination of the cranial capacity as basis for brain mass estimation. The usual method of packing the cranial cavity with small rounded particles as lead shot, mustard seed or millet grain, and other comparable materials is limited by the need to condense the fill both in the skull and in the measuring cylinder in the same way. The method closing the foramina before packing the cranial cavity may also influence the result. A good standardized practice is needed to achieve a tolerable accuracy with this volumetric approach, and it may be combined with a weighing procedure (Smith et al. 1995). Measurements obtained by the packing method may surpass the volume of artificial endocasts (Martin 1990, Table 8.5: 10–11% in the Eocene lemuroid primate Adapis parisiensis). Estimation of endocranial volume by double graphic integration derived from X-ray pictures may result in substantially diverging values (Martin 1990, Table 8.5: 10–12% less than the artificial endocast in Adapis parisiensis). An assessment of cranial capacity in fossil primates also may be done by the use of linear cranial dimensions (Martin 1990) (Table 19.10).

### Table 19.10
**Estimation of cranial capacity (CrC) (cm³) by linear dimensions in primates** (Martin 1990: >Table 8.10)

<table>
<thead>
<tr>
<th>Dimension (mm)</th>
<th>r</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Braincase width (CW)</td>
<td>0.99</td>
<td>log CrC = 3.24 log CW – 3.75</td>
</tr>
<tr>
<td>Braincase height (CH)</td>
<td>0.98</td>
<td>log CrC = 2.91 log CH – 2.91</td>
</tr>
<tr>
<td>Braincase length (CL)</td>
<td>0.98</td>
<td>log CrC = 3.28 log CL – 4.37</td>
</tr>
<tr>
<td>Sum CW + CH + CL (SU)</td>
<td>0.995</td>
<td>log CrC = 3.12 log SU – 5.18</td>
</tr>
<tr>
<td>Product CW x CH x CL (PR)</td>
<td>0.995</td>
<td>log CrC = 1.02 log PR – 3.54</td>
</tr>
</tbody>
</table>

The formulas are based on the major axis, the cranial capacities were determined by packing the cranial cavity with sintered glass particles.

### 19.5 Calculating relative brain size

There are four main concepts of scaling relative brain size in primates (McHenry 1976). The constant of cephalization (CC) (Hemmer 1971) was developed on the base of a common mammalian intraspecific allometric exponent around 0.23
[\text{BrM} = \text{CC} \times \text{BM}^{0.23}; \text{BrM} = \text{brain mass (g)}, \text{BM} = \text{body mass (g)}]. \text{The index of progression (IP) (Bauchot and Stephan 1966, 1969) gives the ratio of actual brain mass to brain mass predicted on the basis of an interspecific basal insectivore allometric equation with an exponent of 0.63. [The index of cranial capacity (ICC) as introduced by Martin (1990) follows just the same lines as the index of progression.]} \text{The encephalization quotient (EQ) (Jerison 1973) is also the ratio of actual to predicted brain mass, the latter based on an interspecific mammalian allometry (predicted brain mass} = 0.12 \text{BM}^{2/3}). \text{The extra neurons count (Nc) (Jerison 1973) “is a numerical measure of progressiveness in brain development beyond the level required by increasing body size” (Jerison 1973). The results of the CC and Nc methods perfectly correlate on the one hand as do the results of the IP and EQ methods on the other (Figure 19.3). This reduces the issues on principle to the choice of either the intraspecific (CC and Nc methods) or the interspecific (IP and EQ) type of brain to body mass allometries.}

\text{EQ has been widely used in the last three decades by the overwhelming majority of authors (Hartwig-Scherer 1993; Kappelman 1996; Ruff et al. 1997; McHenry and Coffing 2000). The results of EQ calculations vary depending on which of the different equations is selected (Kappelman 1996) [allometric exponents varying from 0.60 (Old World simian EQ: Martin 1990) or 0.67 (Jerison 1973) close to 0.75 (0.74–0.76) (Martin 1990; Hartwig-Scherer 1993; Ruff et al. 1997; McHenry and Coffing 2000)]. Nevertheless, some writers have begun to feel uncomfortable about the EQ method (McHenry 1988; Kappelman 1996; Arsuaga et al. 1999; Rightmire 2004). A negative allometric relationship between body size and EQ was raised, but the reasons were assumed to be unclear (Kappelman 1996). EQ being a function of body mass predicted for individuals using an interspecific equation, the comparison of EQ values determined for fossils was considered to be misleading (Rightmire 2004). Similarly, the EQ was not felt to be meaningful when closely related species with widely differing body mass are compared (Arsuaga et al. 1999). Curious EQ results like a position of \textit{Miopithecus talapoin} above all nonhuman hominoids, the gorilla ranging below all cercopithecines (Hemmer 1971: Table 2), or \textit{Cebus albifrons} lying between \textit{Homo erectus} and \textit{Homo sapiens} (Hartwig-Scherer 1993: Fig. 6) also indicate that this procedure may be seriously biologically inadequate. All such problems disappear when the CC intraspecific approach is followed instead of the EQ interspecific method (Hemmer 1971). As a by-effect of the low-intraspecific allometric exponent, the influence of differences in body mass predictions on EQ calculations (Conroy 1987) is less profound with the CC method.}
Bivariate log–log plots (lines: least-squares regressions) to demonstrate the mutual relationship of the main methods to scale relative brain size (data from McHenry 1976: Table 2; hominoid primates). (a) Extra neurons count (Nc) (Jerison 1973) against constant of cephalization (CC) (Hemmer 1971), $r = 0.997$. (b) Encephalization quotient (EQ) (Jerison 1973) against constant of cephalization (CC), $r = 0.971$. (c) Encephalization quotient (EQ) against index of progression (IP) (Bauchot and Stephan 1966), $r = 0.998$. 

\[ N_c \sim 10^r \]

\[ EQ \sim 10^r \]
19.6 Life history data estimations in fossil hominids

The final approach to life history data estimation in fossil hominoids, especially in fossil hominids, is a story of reliability of the prediction equations to be used. There is a wide variability in the results of body mass estimation obtained on the basis of different species samples used to create the equations, on the basis of different dimensions, and using different line-fitting techniques (Table 19.11). The availability of several parts of a single skeleton, as e.g., in the female *Australopithecus afarensis* AL 288-1 specimen, allows for many independent estimates, that group together centrally to resemble a normal distribution, to give a clear and conclusive view of the most probable body mass of that individual, and to allow easy recognition of outsider values (Figure 19.4). In the AL 288-1 case, the peak of the density curve (made up of 54 estimates) is found near 29 kg, the mean and the median range near 30 kg, allowing for a consistent estimation of body mass roughly around 30 kg, as also found in earlier studies (Jungers 1988, 1990; McHenry 1992). The existence of such a key specimen also allows us to judge empirically which predictors may be more useful and which should be excluded from the estimation process for a taxon for which neither humans nor African apes are completely adequate models (Hartwig-Scherer 1993). For AL 288-1, dimensions of the humerus and radius heads and the elbow joint (McHenry 1992) produce clearly estimates that are too small based on general hominoid allometries as does the size of the sacral body (McHenry 1992). On the other hand, very large estimates result from the circumference of the humerus based on *Homo sapiens* allometry and from the circumference of the tibia based on an African ape allometry (Hartwig-Scherer 1993), while the dentition provides an outsider value (Jungers 1988).

The availability of well-founded body mass estimates based on postcranial predictors allows us to determine which cranial predictors compare in reliability with them for *Australopithecus* and early *Homo* as well. Orbital height was extracted as the cranial variable which produces body mass estimates that are most in line with postcranially generated estimates (Aiello and Wood 1994). This may be supplemented by estimates based on a computer digitizing measurement of the orbital area (Kappelman 1996).

Published body mass estimates based on these cranial dimensions were used together with most relevant postcranially predicted data to extract rounded mean body mass for *Australopithecus* and *Homo* paleopopulations and to calculate in each case the CC with rounded mean brain mass converted from mean cranial capacity (Tables 19.12 and 19.13). These brain mass and CC data were then used as the source material for the estimation of life history traits. The evaluation of the reliability of the life history trait predictor equations (Tables 19.3 and
Table 19.11
Selected examples that characterize the broad variability of body mass estimates (kg) of fossil hominids derived at the base of different predicting equations

<table>
<thead>
<tr>
<th>Species and specimen</th>
<th>Body mass predicted by dental dimensions</th>
<th>Body mass predicted by cranial dimensions</th>
<th>Body mass predicted by postcranial dimensions</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Australopithecus (Paranthropus) boisei</em> KNM-ER 406</td>
<td>66.6-64.4 – 59.8-60.1, 57.6-58.8 – 61.3-60.7, 92.3-96.9 – 85.8-85.4, 69.8</td>
<td>84.0-88.4-86.4 – 57.0-57.4-57.3, 50.3 – 48.7 – 57.1-58.2-58.0 – 51.3-51.4-51.4, 66.3-68.4-68.2 – 48.3-48.4-48.3, 54.9-55.9-55.7 – 43.2-43.2-43.2, 55.8-57.1-56.9 – 42.4-42.3-42.3, 46, 58.4, 86.9 – 106.5, 45 – 41, 48 – 36, 50 – 35, 43</td>
<td>Aiello and Wood (1994), Kappelman (1996)</td>
<td></td>
</tr>
<tr>
<td><em>Homo sp. (rudolfensis)</em> [Australopithecus (Paranthropus) boisei ?] KNM-ER 1481</td>
<td>80.4-77.4 – 70.5-70.6, 50.7-51.9 – 54.2-53.9, 95.2-99.8 – 88.1-87.7, 45.6</td>
<td></td>
<td>McHenry (1988), McHenry (1992), Ruff and Walker (1993), Hartwig-Scherer (1993)</td>
<td></td>
</tr>
</tbody>
</table>

Mass estimates separated by - : prediction based on the same dimension and the same species sample, but different line fitting technique; mass estimates separated by – : prediction based on the same dimension, but different species samples; mass estimates separated by , : prediction based on different dimensions.
19.4 was done empirically based on the PE of the estimation of *Homo sapiens* [PE = (observed – predicted)/predicted × 100]. Only equations with a PE ≤ 5% were considered to be useful for the life history estimation in fossil hominids. It must never be forgotten that most of the actually observed primate life history dimensions are subject to enormous variability produced by diverse
Table 19.12
Estimates of body mass, brain mass, and constant of cephalization in *Sahelanthropus* and *Australopithecus*

<table>
<thead>
<tr>
<th>Taxon</th>
<th>Sex</th>
<th>Body mass, cranial estimate (kg)</th>
<th>Body mass, postcranial estimate (kg)</th>
<th>Retained body mass (kg)</th>
<th>Cranial capacity (cm³)</th>
<th>Brain mass estimate (g)</th>
<th>Retained brain mass (g)</th>
<th>Constant of cephalization</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Sahelanthropus tchadensis</em></td>
<td>male?</td>
<td>58¹</td>
<td></td>
<td>c. 58</td>
<td>c. 320–380¹</td>
<td>302–357</td>
<td>c. 300–360</td>
<td>c. 24–29</td>
</tr>
<tr>
<td><em>Australopithecus afarensis</em></td>
<td>Male</td>
<td></td>
<td></td>
<td>c. 45</td>
<td>c. 30</td>
<td>434</td>
<td>407</td>
<td>c. 405</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td></td>
<td></td>
<td>c. 40</td>
<td>401, 454</td>
<td>423</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Both sexes</td>
<td></td>
<td></td>
<td>c. 40</td>
<td>c. 35</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Australopithecus africanus</em></td>
<td>Male</td>
<td>30–68</td>
<td>41–41</td>
<td>c. 40</td>
<td>c. 30</td>
<td>428, 485</td>
<td>401, 454</td>
<td>c. 430</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td></td>
<td></td>
<td>c. 30</td>
<td>425</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Both sexes</td>
<td>30, 22</td>
<td>33–58</td>
<td>c. 35</td>
<td>423</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>A. (Paranthropus aethiopicus)</em></td>
<td>Male</td>
<td>38</td>
<td></td>
<td>c. 40</td>
<td>410</td>
<td>385</td>
<td>c. 385</td>
<td>c. 34</td>
</tr>
<tr>
<td><em>A. (Paranthropus boisei)</em></td>
<td>Male</td>
<td>58, 39</td>
<td>49–51</td>
<td>c. 50</td>
<td>510, 530</td>
<td>476, 495</td>
<td>487</td>
<td>c. 485</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>70, 58</td>
<td>50</td>
<td>c. 50</td>
<td>500</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Both sexes</td>
<td>45–51</td>
<td>45–51</td>
<td>c. 40–45</td>
<td>521</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>A. (Paranthropus robustus)</em></td>
<td>Male</td>
<td>47</td>
<td>40–56</td>
<td>c. 40</td>
<td>530</td>
<td>495</td>
<td>c. 495</td>
<td>c. 44</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td></td>
<td></td>
<td>c. 30</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Both sexes</td>
<td>47</td>
<td>37–56</td>
<td>c. 35–40</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
ecological factors. Any prediction for extinct populations will be subject to this variability too.

Nevertheless, the results then obtained confirm, correct, and supplement earlier [Hemmer (1974) on the age at sexual maturity, at first breeding, and at teeth eruption in *Australopithecus*] and later calculations [Smith and Tompkins (1995) on first permanent molar eruption]. At the same time, they are supported by results of osteobiographic techniques. *Australopithecus* (*afarensis, africanus, robustus*) and early *Homo* specimens, aged at something like 3–3.5 years on the basis of incisor crowns with little or no root development, had first permanent molars coming into occlusion (Bromage 1990) just as predicted with first lower molar eruption at around 3 years (Table 19.14). The estimation of life history data confirms the fundamental nature of the australopithecines as progressive apes, not as humans, the early *Australopithecus afraransis* ranging within the modern African ape life history dimensions.

Some important new fossil taxa discovered in the last years are included in the selected samples of life history data presentation (Table 19.14). The interpretation of the Late Miocene *Sahelanthropus tchadensis* (Brunet et al. 2002), as a taxon possibly related to the hominid stem species near the splitting of chimpanzee and hominid lines (Brunet et al. 2002; Wood 2002), may be supported by the estimation of life history dimensions that indicate an earlier evolutionary stage than all *Australopithecus* and *Homo* paleopopulations at the one hand and than the chimpanzee on the other. The Lower Pleistocene (Upper Villafranchian) *Homo* paleopopulation of Dmanisi (Georgia) (*Homo georgicus* Gabounia et al., 2002) has been interpreted as more closely related to the Late Pliocene *habilis and rudolfensis* than to the Lower Pleistocene *ergaster and erectus* (Gabunia et al. 2000; Gabounia et al. 2002; Vekua et al. 2002). The life history estimates clearly support this view. The so-called *Homo floresiensis* (Brown et al. 2004) is not integrated here, as this specimen obviously neither represents a new species nor an enigmatic branch of hominid evolution, but is just a classic microcephalic *Homo sapiens* individual, sharing the characteristic syndrome of pygmy size, very small (chimpanzee sized) brain, and considerably aberrant skull.

---

Original body mass values rounded to the next kg, retained body mass to the next 5 kg. Brain mass estimated on the basis of the formula given by Martin (1990) (Table 19.9a), retained brain mass rounded to the next 5 g.

*Aiello and Wood (1994, mean of LSR and RMA estimates).
**Kappelman (1996).

1Basic dimensions by Brunet et al. (2002): orbital height used to estimate body mass with equations (Table 19.6) published by Aiello and Wood (1994) and Kappelman (1996).
### Table 19.13

Estimates of body mass, brain mass, and constant of cephalization in Pliocene and Pleistocene *Homo*

<table>
<thead>
<tr>
<th>Taxon</th>
<th>Sex or specimen</th>
<th>Body mass, cranial estimate (kg)</th>
<th>Body mass, postcranial estimate (kg)</th>
<th>Retained body mass (kg)</th>
<th>Cranial capacity (cm³)</th>
<th>Brain mass estimate (g)</th>
<th>Retained brain mass (g)</th>
<th>Constant of cephalization</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Homo</em> sp. <em>habilis</em>, latest Pliocene</td>
<td>Male</td>
<td>34, 26</td>
<td>30, 35</td>
<td>37</td>
<td>594, 509</td>
<td>553,476</td>
<td>c. 515</td>
<td>c. 48</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>30, 35</td>
<td></td>
<td></td>
<td>c. 35</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Both sexes</td>
<td>30, 35</td>
<td></td>
<td></td>
<td>c. 30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>c. 35</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Homo</em> sp. <em>rudolfensis</em>, latest Pliocene</td>
<td>Both sexes</td>
<td>51</td>
<td>47</td>
<td>51–60</td>
<td>752</td>
<td>697</td>
<td>c. 695</td>
<td>c. 58</td>
</tr>
<tr>
<td><em>Homo</em> sp. <em>georgicus</em>, Lower Pleistocene</td>
<td>Female</td>
<td>30</td>
<td></td>
<td></td>
<td>c. 30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Both sexes</td>
<td></td>
<td></td>
<td></td>
<td>c. 30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Homo</em> sp. <em>ergaster</em>, Lower Pleistocene</td>
<td>Both sexes</td>
<td>58–66</td>
<td>57–60</td>
<td>56–66</td>
<td>c. 60</td>
<td>804–909</td>
<td>805</td>
<td>c. 64</td>
</tr>
<tr>
<td><strong>Homo sp.</strong></td>
<td><strong>Zhoukoudian XI</strong></td>
<td><strong>Zhoukoudian XII</strong></td>
<td><strong>Sangiran 17</strong></td>
<td>Both sexes</td>
<td><strong>c. 60</strong></td>
<td><strong>c. 935</strong></td>
<td><strong>c. 74</strong></td>
<td></td>
</tr>
<tr>
<td>-------------</td>
<td>-------------------</td>
<td>-------------------</td>
<td>----------------</td>
<td>------------</td>
<td>----------</td>
<td>-----------</td>
<td>----------</td>
<td></td>
</tr>
<tr>
<td><strong>erectus,</strong> Early Middle Pleistocene</td>
<td>52</td>
<td>1015</td>
<td>935</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>66</td>
<td>1030</td>
<td>949</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>58</td>
<td>1004</td>
<td>925</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>76</td>
<td>1030</td>
<td>925</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>archaic Homo sapiens,</strong> Middle Pleistocene, Europe</td>
<td>35′</td>
<td>1166′</td>
<td>1071</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>94″</td>
<td>1390”</td>
<td>1273</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>35′</td>
<td>1100′</td>
<td>1129</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>52′</td>
<td>1230’</td>
<td>1230</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Both sexes</td>
<td><strong>c. 55</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1120</td>
<td>1230</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1140</td>
<td>1120</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Late archaic Homo sapiens</strong> (Neanderthals)</td>
<td>Male</td>
<td>81–100/92</td>
<td>1565</td>
<td>1430</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Both sexes</td>
<td><strong>c. 90</strong></td>
<td></td>
<td></td>
<td></td>
<td><strong>c. 1430</strong></td>
<td><strong>c. 104</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>76</td>
<td><strong>c. 75</strong></td>
<td>1489****</td>
<td>1370</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>c. 1370</strong></td>
<td><strong>c. 104</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Original body mass values rounded to the next kg, retained body mass to the next 5 kg. Brain mass estimated on the basis of the formula given by Martin (1990) (<sup>2</sup> Table 19.9a), retained brain mass rounded to the next 5 g.

* Aiello and Wood (1994, mean of LSR and RMA estimates).
** Kappelman (1996).
**** Ruff et al. (1997).
' Rightmire (2004).
# Arsuaga et al. (1999).

<sup>1</sup> basic dimensions of skull D 2700 by Vekua et al. (2002): orbital height used to estimate body mass with equations (<sup>2</sup> Table 19.6) published by Aiello and Wood (1994) and Kappelman (1996); additional cranial capacities of D 2280 and D 2282 by Gabunia et al. (2000).
allometries (Hemmer 1967). Perhaps this fossil may be interesting as an example of the survival of handicapped people in a Paleolithic culture. It may also help to understand the dysregulation leading to microcephalic development as a possible key to an intrinsic relationship of brain, maturation, and life history.

Acknowledgments

The author thanks Prof. Dr. Winfried Henke for various valuable supports.
References


Estimation of basic life history data of fossil hominoids


Abstract

The study of population genetics can contribute to paleoanthropological research in three ways. First, the analysis of genetic variation in living human populations can provide information about past events, including the time and place of recent common ancestors, which when considered over many loci can give us clues regarding patterns of human evolution. Recent work in this area supports a view of several dispersions out of Africa, with the later two corresponding roughly to the appearance of the morphospecies *Homo heidelbergensis* and modern *H. sapiens*, but in both cases showing evidence of dispersions with admixture, rather than dispersions with replacement. Second, studies of ancient DNA, such as those conducted to date on Neandertal fossils, can potentially give us insight into ancient genetic variation and relationships among populations or taxa. Evidence to date suggests that Neandertals may have contributed little in overall ancestry to living humans, but it is still not clear whether they contributed nothing genetically. Third, mathematical models and simulations based on population genetic theory can help us formulate and refine hypotheses regarding human evolution by providing an idea of what patterns should be expected in the fossil record under alternative evolutionary scenarios.

20.1 Introduction

The use of genetic data and genetic inferences in paleoanthropological research has become increasingly accepted in recent years. This was not always the case. Sarich and Wilson (1967) proposed a method of “molecular dating” that when applied to data from extant hominoid species led to the then-radical conclusion that the African ape and human lineages had diverged only 4–5 Ma, rather than the more accepted 15+ million-year date based on many interpretations of the fossil record available at that time. Over time, the late-divergence hypothesis became accepted by the majority of the paleoanthropological community because of both additional fossil evidence as well as a growing acceptance of molecular
methods (Lewin 1987; Tattersall 1995). Since the mid-1980s, genetic data and inferences have become part of the effort to unravel the origins of modern humans. Over the years, numerous books and articles have reviewed the subject of genetics and modern human origins. My purpose here is not to provide a detailed overview of previous studies and reviews, but rather to highlight the major uses and possible future directions of population genetic data and models in the search for modern human origins.

20.2 Models of modern human origins

The modern human origins debate is concerned with the evolutionary relationships between “archaic” and “modern” humans over the past 200,000 years. Although it is clear on a very general level that archaic humans gave rise to anatomically modern humans, the specifics of this transition make up the modern human origins debate. A certain amount of the debate over modern human origins may lie in differing interpretations of the evolutionary models that have been proposed to explain this transition. Although it is quite typical to see this debate phrased in terms of a conflict between a “recent African origin” and “multiregional evolution,” in reality the situation has been more complex, in part due to the fact that there are several variants of each model, sometimes overlapping in terms of underlying predictions. Elsewhere (Relethford 2001c), I attempted to condense much of this debate by focusing on two basic questions: (1) what was the mode of this transition? and (2) what is the location and timing of this transition? The joint answer to both of these questions determines the specific nature and predictions of various modern human origin models.

The mode of transition refers to whether modern humans arose through speciation and replacement of archaics outside the region of transition or through evolution within a species across most or all of the Old World. In others words, the debate here is about cladogenesis versus anagenesis. Although this distinction is clear in terms of the contrast between the recent African origin and multiregional evolution models, the question of the location and timing of the transition is fuzzier. Here, there are two different views: one arguing that the transition took place in a single area (Africa) at a specific time, and the other arguing that there was no single time or place associated with the origin of modern humans. Rather, the transition from archaics to moderns occurred piecemeal across the Old World, with such changes coalescing throughout the entire species due to gene flow. Although much of the literature supporting multiregional evolution takes this latter view, it is important to realize that there are advocates of multiregional evolution (as a general process) that argue
for a single and recent African origin of modern humans (Wolpoff et al. 1994; Relethford 2001c).

### 20.2.1 Models

I have argued elsewhere that consideration of the issues of mode and timing/location leads to three general classes of models (Relethford 2001c). The first is the African Replacement Model, which argues that modern humans arose through cladogenesis as a new species (*Homo sapiens*) in Africa roughly 150 to 200 ka or so. Over time, this new species expanded out of Africa and eventually replaced all preexisting archaic populations outside of Africa. Under this model, various non-African archaics, such as the Neandertals of Europe, became extinct and did not contribute genetically to living humans because they were, by definition, a separate species. This model links the mode of cladogenesis with the idea of a single origin point.

The second model is the Regional Coalescence Model. This model is usually associated with one commonly used definition of multiregional evolution. Here, the transition from archaics to moderns took place within a single evolving lineage (anagenesis) and there was no single time or place associated with this transition. Proponents of this view argue that the boundary between “archaic” and “modern” implies a lack of a clear point of transition. Instead, changes occurred over time in different geographic regions, which ultimately were shared across the range of the species through gene flow (although it is not necessary under this model that *all* populations ultimately contributed ancestry to modern humans).

The third model is the Primary African Origin Model. This model shares the idea of a species-wide transition with the Regional Coalescence Model and the idea of a single point of origin (Africa) with the African Replacement Model. Here, the transition from archaic to modern humans took place in Africa roughly 150 to 200 ka, followed by an expansion of some populations out of Africa by 100 ka. However, unlike the African Replacement Model, this model allows admixture between expanding modern populations and the preexisting archaic populations outside of Africa. Different variants of this model emphasize different forms of this admixture, including isolation by distance effects, long-range migration, demic diffusion, and various combinations of all of the above. The Primary African Origin Model shares not only an explanation of mode of evolution with the Regional Coalescence Model but also the idea of a specific recent African origin with the African Replacement Model. Because the Primary African Origin Model invokes genetic ancestry from more than one region, it is by broad
definition a multiregional model. However, because of its proposed single and recent origin of anatomical modernity, it is also by broad definition a recent African origin model. Some examples of models which I feel fall into the category of “Primary African Origin” models are Bräuer’s (1992) “Afro-European hybridization” model, Smith et al.’s (1989) “assimilation” model, Relethford’s (2001c) “mostly out of Africa” model, and Eswaran’s (2002) “diffusion wave” model.

20.2.2 Fossil record

Before considering the relevance of genetics to testing of these models, it is necessary to briefly consider the current status of the fossil record on the timing and appearance of anatomically modern humans. Although not everyone agrees, there appears to be growing consensus that the earliest evidence of modern human anatomy is in Africa dating back more than 100 ka, perhaps as far back as 195 ka at Omo (McDougall et al. 2005). Recent discoveries from the Middle Awash show evidence of a transition between archaic and modern morphology almost 160 ka (White et al. 2003). If the finding of an earlier presence of anatomically modern humans in Africa persists, then it seems likely that the Regional Coalescence Model of modern human origins will be rejected. The fossil evidence, along with genetic evidence pointing to a central importance of Africa in modern human evolution, supports a recent African origin, although this by no means automatically supports a speciation and subsequent replacement. Although this point gets glossed over in many treatments of the modern human origins debate, a recent African origin is not incompatible with some versions of multiregional evolution, specifically the Primary African Origin model (Relethford 2001c). In terms of the two debate issues discussed earlier (mode of transition and timing/location of transition), the latter issue seems settled for many anthropologists—modern humans began, as an anatomical form, in Africa about 150 to 200 ka. What remains contentious is the genetic interaction, if any, between modern humans dispersing out of Africa and the preexisting archaic humans outside of Africa. Thus, the debate at present is primarily between replacement and admixture. A further complication is the distinct possibility that both replacement and admixture occurred, each in different areas outside of Africa.

20.2.3 Population genetics and modern human origins

In broad terms, the study of population genetics can be applied to the study of modern human origins in three ways: (1) analysis of patterns of genetic variation in living human populations, (2) analysis of patterns of ancient genetic
variation, specifically ancient mitochondrial DNA (mtDNA), and (3) application of models of population genetics to the fossil record. Of these, the first two are widely known and discussed, whereas the important contributions of population genetics models for interpreting the fossil record have received less attention, but it is likely to be an important concern of future research in modern human origins.

20.3 Patterns of genetic variation in living human populations

Anthropologists and human biologists have been collecting information on patterns of biological variation within and between living human populations for several centuries, starting with anthropometrics and other physical measures, followed by blood types in the early twentieth century, and blood proteins and enzymes in the latter half of the twentieth century. Over the past two decades, newer genetic markers brought about through an ongoing revolution in molecular genetics have allowed us even closer glimpses and genotypes and underlying genetic code. This massive and everincreasing store of data on contemporary human genetic variation has definite implications for paleoanthropological research. The rationale here is that whatever happened in the past, these events have left a genetic signature on patterns of contemporary genetic variation. Our current patterns of genetic diversity are in essence “reflections of our past” (Relethford 2001c, 2003). The trick is being able to discern ancient population dynamics from current patterns of genetic variation in such a way that specific hypotheses can be tested. In some cases, the immediately obvious inferences made from contemporary genetic variation have turned out to be more complex than once thought, or often reflecting several possible past histories.

20.3.1 Gene trees

The molecular revolution has allowed detailed glimpses into the mutational history of specific genes and DNA sequences. Reconstruction of gene genealogies (gene trees) allows us the ability to identify the most recent common ancestor (MRCA) of any set of sequences from contemporary populations. If our sampling is geographically diverse, identification of the most recent common ancestor provides information on all of humanity, which in turn can have implications for modern human origins. Such analyses can then tell us something about the time back to the most recent common ancestor (TMRCA) and the place of the most recent common ancestor (PMRCA) (Takahata et al. 2001).
The best known example of this type of work resulted from early analyses of mtDNA variation in living humans. Mitochondrial DNA is a useful trait for genealogical analysis because it is inherited strictly through the maternal line and does not recombine (Stoneking 1993). The relatively high mutation rate of mtDNA means it is particularly useful for more “recent” evolutionary events. The application of mtDNA variation to the debate over modern human origins came with the study of Cann et al. (1987) which showed that the most recent common mtDNA ancestor lived in Africa roughly 200 ka. These findings were later verified by other studies (Vigilant et al. 1991; Penny et al. 1995). The significance of these findings lay in the fact that the most recent common ancestor lived well after the initial dispersion of Homo out of Africa. The date of 200 ka (even given a wide confidence interval) led many to reject any sort of multiregional interpretation of modern human origins on the grounds that multiregional evolution, with its emphasis on a single evolving lineage of Homo, required common ancestry back to 2 Ma. One problem, later described in many reviews (Relethford 2001c; Klein and Takahata 2002), is that the mtDNA gene tree described the history of that particular trait alone, and that the history of a gene or DNA sequence may not necessarily reflect the history of a population. It has since been realized that a more accurate reconstruction of modern human origins requires synthesis of a number of different gene trees, including those available from mtDNA, Y-chromosome DNA, and segments of nuclear DNA that show little if any recombination.

Several studies have examined patterns of common ancestry based on multiple gene trees. Takahata et al. (2001) examined data from autosomal, X-chromosome, Y-chromosome, and mitochondrial sequences. Nine out of 10 of these sequences showed the most recent common ancestor in Africa, and the 10th sequence was Asian in origin. They argued that the results favored a single African origin but did not rule out some non-African ancestry under the condition that the founding population in Africa was much larger than outside of Africa. This conclusion is interesting because other genetic data support a larger long-term African population size (Relethford and Harpending 1994; Relethford and Jorde 1999). The dominant genetic signature from Africa, combined with evidence of some non-African ancestry, is best explained by a primary African origin model.

Another multivariate approach to gene tree analysis has been undertaken by Templeton (2002, 2004), who used a complex method known as nested clade phylogeographic analysis which can identify both unique historical events, such as population dispersions, and ongoing recurrent events, such as gene flow. This method was applied to data on mtDNA, Y-chromosome DNA, X-chromosome DNA, and autosomal DNA haplotypes. The comparison of results across these difference sequences showed that there has been more than one dispersal out of
Africa (Templeton 2004). The mtDNA and Y-chromosome DNA sequences indicated a fairly recent dispersion with an estimated age of 80–150 ka. However, other sequences showed another earlier dispersal out of Africa occurred roughly 420–840 ka. Combined with the fossil evidence for a dispersal of \textit{Homo} out of Africa roughly 1.8 Ma, Templeton (2002) therefore argued that there have been three distinct African dispersals. Although the confidence intervals are broad, the latter two African dispersals coincide roughly with morphological changes in \textit{Homo}: the increase in brain size roughly 700 ka (Relethford 2001c) and the emergence of anatomically modern human crania out of Africa 100 ka.

Multiple dispersions out of Africa can fit either a replacement model or an admixture model. In the case of the replacement model, each dispersal might correlate with the rise of a new species in Africa: \textit{H. erectus} followed by \textit{H. heidelbergensis} followed by \textit{H. sapiens}. One could argue that in each case a new species arose in Africa and then dispersed across the Old World to replace the preexisting species. In the case of an admixture model, much the same thing could have occurred, but with new morphological changes spreading through gene flow, rather than replacement, outside of Africa.

Templeton’s nested clade phylogeographic analyses provide some resolution to the replacement/admixture debate by noting different predicted patterns of haplotype variation under conditions of range expansion and those of recurrent gene flow and isolation by distance. Templeton (2002) concluded that there is ample evidence of recurrent gene flow between geographic regions throughout the past 2 Myr, suggesting that the two dispersals out of Africa following the initial dispersion of \textit{H. erectus} resulted in interbreeding rather than replacement. If true, then the evolution of the genus \textit{Homo} has been characterized by a series of expansions out of Africa, with each subsequent expansion occurring on top of previous ones, in addition to some level of interregional gene flow over time. The appearance of new morphological variants, such as archaic and modern humans, would thus represent an evolutionary shift within a single lineage rather than a series of speciation events. It is less clear, however, how to distinguish these possible scenarios in the fossil record. Depending on the relative speed of dispersal and replacement versus admixture, the two models might produce essentially the same pattern of morphological change within \textit{Homo}—early humans (\textit{H. erectus}) followed by archaic humans followed by modern humans.

### 20.3.2 Genetic diversity

Another significant finding of population genetics research on living populations has been the discovery that sub-Saharan African populations show greater
diversity than do populations in other geographic regions. This pattern of higher levels of variation has been found in mtDNA (Cann et al. 1987; Vigilant et al. 1991), microsatellite DNA markers (Jorde et al. 1995), Alu insertion markers (Watkins et al. 2001), craniometrics (Relethford and Harpending 1994), and skin color (Relethford 2000).

The evolutionary significance of higher sub-Saharan African diversity, and its implications for the question of modern human origins, has been debated. One explanation is that diversity is a reflection of the age of a population; the longer a population has been around, the more mutations have accumulated, and hence the older it is. Under this model, as a daughter population splits off from a parental population to form a new founding group, the small size of the founding group results in genetic drift such that the level of diversity is “reset” and future mutations then accumulate since the time of founding. Given the assumption of diversity correlating with age, it has been argued that higher levels of sub-Saharan African diversity thereby support a replacement model. If modern humans arose first in Africa, and non-African populations formed as founder groups later in time, then we would expect to see exactly the same pattern as we find today—higher levels of genetic diversity in sub-Saharan Africa. Although this finding does at first glance support an African origin of modern humans, it does not necessarily rule out admixture outside of Africa. In addition, we also need to be cautious when making inferences about age from diversity because the underlying assumptions about magnitude and duration of a bottleneck effect may be too unrealistic to apply to the evolution of modern humanity (Rogers and Jorde 1995; Relethford 2001c).

Furthermore, there are other factors that influence diversity that are more realistic, primarily gene flow and population size. As shown in analyses of craniometric data (Relethford and Harpending 1994) and microsatellite DNA markers (Relethford and Jorde 1999), higher levels of phenotypic/genetic diversity in sub-Saharan Africa are likely due to larger population size in Africa. Because of the relationship between population size and genetic drift, smaller populations will experience more genetic drift and hence have lower levels of diversity. A larger African population size is consistent with likely prehistoric population density, ecology, and overall land mass (Hassan 1981; Thorne et al. 1993).

It is important to note, however, that a larger African population size is possible under virtually any model of modern human origins, and the fact that genetic evidence points to this finding does not verify or rule out either replacement or admixture. On the other hand, it does point again to the importance of Africa in the genetic structure of living humans. A larger long-term African
population size means that, evolutionarily, the majority of our ancestors in recent evolutionary history are likely to have lived in Africa. However, it is not clear whether all of our recent ancestors are African, as required by a speciation and replacement model. As I suggest elsewhere (Relethford 2003), the question boils down to where our ancestors lived some 150 to 200 ka. Under a replacement model, all of them were African, whereas all other models suggest that some of them were African. The higher levels of African genetic diversity, likely reflecting larger African population size, agree with other genetic and fossil evidence of an African origin for anatomically modern humans, but it does not distinguish between replacement and admixture.

Similarly, geographic patterns in genetic diversity can be explained under either a replacement or admixture model. Some genetic studies have found a pattern of diversity where non-African diversity is a subset of African diversity in terms of haplotypes (Tishkoff et al. 1996; Watkins et al. 2001). Although this pattern fits nicely with an African origin and subsequent dispersion and replacement, it can also be accommodated with an admixture model with larger African population size. In simplest terms, lower non-African population sizes throughout much of human evolution would have caused greater genetic drift outside of Africa, which in turn could lead to non-African diversity reflecting a subset of African diversity due to loss of haplotypes under genetic drift. Again, this could have happened under either a replacement or admixture scenario depending on particular parameter values.

### 20.3.3 Population genetics and coevolution of host and parasite

A new type of genetic analysis that can provide insight into human evolution is the study of the genetic history of human parasites. One recent example is the study of Reed et al. (2004) on the evolutionary history of the modern human head louse, *Pediculus humanus*. This species has two lineages, one globally distributed and the other found only in the New World. Genetic analysis suggests that these two lineages separated roughly 1.2 Ma. The presence of both louse lineages in living humans is best explained by a host switch later in time after this divergence, which in turn suggests close physical contact between *H. erectus* or archaic humans and modern humans. These results do not resolve the debate over replacement versus admixture because the physical contact between previously divergent human populations might have been nonsexual, or if sexual it might not necessarily have resulted in fertile offspring.
20.3.4 What does the genetic evidence tell us?

Patterns of genetic variation within and between living populations can have a number of different causes that do not distinguish clearly between certain models of modern human origins. This problem of indeterminacy also applies to other observations from genetic data, such as patterns of genetic distance between populations and observations of prehistoric population expansions (Relethford 2001c). Taken as a whole, the genetic evidence certainly points to sub-Saharan Africa as having had a major effect on shaping contemporary patterns of human genetic diversity probably due to an earlier appearance of anatomically modern humans (as seen in the fossil record) combined with a larger African population size. Again, the difficulty is resolving whether this African origin was followed by replacement or admixture. Many classic models of genetic analysis, many of which are based on an a priori assumption of no gene flow, are unable to resolve this problem, and it will require the development of new methods. To date, the most promising is Templeton’s nested clade phylogeographic analysis method. Given the haplotype data collected to date, his analyses suggest that the most parsimonious model of modern human origins is a set of African dispersals overlain on top of each other, along with some continuing recurrent gene flow. If these results hold up as more genetic data accumulate, then the data are best interpreted as showing admixture outside of Africa, although it remains to be seen how much admixture occurred in any given geographic region.

20.4 Ancient DNA

Another genetic approach to human evolution is the relatively new analysis of ancient DNA sequences. I recall a conversation years ago in graduate school where someone made the statement that many debates in human evolution could be solved if only we had genetic data from the past. Such statements seemed harmless enough at the time, given that the idea of extracting genetic data from fossil remains was essentially science fiction at the time, and we expected that such accomplishments were impossible. We were wrong, but on two counts. First, it is possible under appropriate conditions to sequence ancient DNA. Second, the data are not always quite as clear as we imagined they would be. Although some scientists have taken the results from ancient DNA analysis as unambiguous evidence pertaining to modern human origins, others are still skeptical about exactly what these data show us.
20.4.1 Neandertal DNA

The revolution in the use of ancient DNA analysis in the modern human origins debate came with the sequencing of small sections of mtDNA from the Feldhofer Neandertal (Krings et al. 1997, 1999). In the original analysis, Krings et al. (1997) compared a 378-bp Neandertal sequence with mtDNA sequences from living humans. The results were very clear in showing that the Neandertal mtDNA sequence was quite a bit different; the average difference between Feldhofer and living humans is 27 base substitutions, over three times the average difference between random pairs of living humans. Taking this difference as supporting the idea that Neandertals were a separate species, they further derived an estimate of between 550 and 690 ka as the time when the Neandertals and our own ancestors diverged. Since the initial sequencing of the Feldhofer specimen, additional Neandertal mtDNA sequences of varying length have been extracted from several specimens at different sites (Krings et al. 2000; Ovchinnikov et al. 2000; Schmitz et al. 2002; Serre et al. 2004). Collectively, these sequences cluster together and are quite different from those of living humans (although see Gutierrez et al. 2002 for an alternative analysis).

To date, several findings have generally, although not universally, been regarded as supporting replacement of the Neandertals by anatomically modern humans. One such finding has been the overall magnitude of the difference between the Neandertal DNA sequences and those from living humans. However, the meaning of this difference is less clear, as studies on comparable mtDNA sequences in living chimpanzees show that the difference between two out of three comparisons of chimpanzee subspecies is less than that between Neandertals and living humans (Krings et al. 1999). Based on this finding alone, one could argue that Neandertals were at most a different subspecies from us and not a different species (Relethford 2001c).

Another argument that has been made in favor of the replacement model is the fact that Neandertal mtDNA sequences (from Europe) are no more similar to living Europeans than to living humans from other regions (Krings et al. 1997). That is, Neandertal DNA is essentially equidistant from living humans in all geographic regions, a pattern that is consistent with the view that Neandertals diverged before the geographic dispersal of modern humans, and therefore were a separate species. Elsewhere I have questioned this conclusion, and showed that under conditions of interregional gene flow and given sufficient time, the same equidistant pattern of similarity could result without replacement (Relethford 2001a). Of course, this finding does not mean that the replacement hypothesis
is wrong, but instead that the equidistant pattern of mtDNA similarity does not necessarily reject alternatives.

**20.4.2 Early modern human DNA**

The strongest evidence supporting replacement and the extinction of the Neandertal gene pool is the lack of Neandertal sequences in any of the thousands of mtDNA sequences obtained to date from living humans. Although one explanation for the lack of Neandertal-like sequences in living humans is replacement, another possibility is that Neandertals constitute part of the ancestry of living humans but mtDNA lineages have been lost over time because of genetic drift. Adcock et al. (2001b) obtained mtDNA sequences from a number of prehistoric Australian modern human fossils including the Lake Mungo 3 (LM3) fossil, which might date back over 60,000 years. The mtDNA sequence for LM3 was different from both Neandertals and living humans. One explanation is that the mtDNA lineage in LM3, a fossil that is clearly anatomically human, has been lost due to drift over thousands of years. Although this find does not bear directly on the Neandertal issue, the fact that drift could cause the loss of an mtDNA lineage in one part of the world (Australia) means that it is possible the same thing happened in Europe (Relethford 2001b). Again, however, the observation that something could have happened does not mean that it did happen (see also Cooper et al. 2001 and Adcock et al. 2001a for debate over possible contamination and analysis).

One potential solution to the problem of drift over time is the comparison of Neandertal DNA sequences with those obtained from early modern European fossils rather than with living Europeans. Caramelli et al. (2003) extracted mtDNA sequences from two anatomically modern human fossils from Italy dating to 24 ka. These sequences were quite similar to those from living humans but different from Neandertal sequences. One interpretation of these results is that there has been little change in mtDNA sequences in the modern human lineage over time, and therefore that Neandertals, with much greater sequence differences, were a separate species. One problem with this interpretation is that contamination from living humans could lead to the same result. In an attempt to get around the problem of potential contamination, Serre et al. (2004) have suggested that it is more appropriate to determine if Neandertal-like sequences rather than present-day sequences were present in early modern human fossils. Using Neandertal DNA primers, they showed that there were no Neandertal-like sequences present in five early modern human fossils. Based on these results, the authors also computed that a maximum of 25% Neandertal ancestry into the modern
human gene pool would be compatible with a total lack of Neandertal sequences in living humans. They also noted that the maximum amount of Neandertal ancestry is highly dependent on the specific demographic scenario and the amount of Neandertal ancestry is likely to be even lower, a point also addressed by Currat and Excoffier (2004), who argue that there would have been virtually no Neandertal admixture under a realistic model of range expansion into Europe. It is clear from these two studies that the computations depend critically on the specific demographic scenario used, and further work on incorporating archeological findings needs to be done.

20.4.3 Admixture or replacement of Neandertal DNA?

At present, all available data from ancient DNA analysis point to a relatively low amount of Neandertal ancestry in modern humans, but debate remains as to whether it is zero (as expected under replacement). It may turn out that we will be unable to distinguish between a very low amount of Neandertal ancestry, perhaps on the order of a few percent, and zero ancestry. In either case, we could all agree on the genetic replacement of the Neandertal gene pool for all practical purposes but still disagree on the underlying mechanism of anagenesis versus cladogenesis. It is worth noting in this context that there are some indications from the fossil record that agree with a pattern of relatively rapid (by geological standards) modern admixture into the Neandertal gene pool resulting over time in close to zero Neandertal ancestry in living humans. Wolpoff (1999) presents several examples of traits considered unique to Neandertals that are found in reduced frequencies in the earliest post-Neandertal modern humans. The presence of such traits is perhaps explained by the incorporation of Neandertal genes into an expanding modern human population, particularly given the genetic evidence of a larger population size out of Africa which would “swamp” the genetic contributions of a smaller Neandertal population. Wolpoff also found that these traits had, in many cases, reduced in frequency even further in living Europeans, a pattern suggestive of the cumulative impact of admixture. Additional samples over time and space will be needed to confirm this pattern. We also need to consider the fact that even if it turns out that the Neandertals were replaced in Europe, this does not necessarily mean that the same scenario happened elsewhere in the Old World. Rather than to continue in a debate over replacement versus admixture, we may have to acknowledge the possibility that there was both replacement and admixture events involved in the origin of modern humans and consider evidence, both fossil and genetic, from each major geographic region.
20.5 Application of population genetic models to the fossil record

The final way in which population genetics can contribute to paleoanthropology is in the application of models of population genetics to direct interpretation of the fossil record. Although this approach has not been widely applied, it has great potential in both resolving paleoanthropological debates and potentially uniting geneticists and paleontologists in multidisciplinary studies. Apart from data on contemporary or ancient populations, the field of population genetics has much to offer in terms of helping to frame models and hypotheses regarding the fossil record. Here, I present several examples of such an approach.

20.5.1 Per-generation gene flow and accumulated ancestry

Gene flow generally acts to reduce genetic differences between populations. The degree to which populations are distinct from one another depends in part on the magnitude of gene flow between them, be it local gene flow mediated by isolation by distance or long-distance dispersal. Expectations of the relationship between fossil samples separated by time and space are dependent on rates of gene flow. However, the concept of per-generation gene flow is sometimes confused with the idea of accumulated ancestry over many generations. One example is the prediction sometimes made that multiregional evolution requires greater similarity between fossil samples in the same region over time (e.g., European Neandertals and living Europeans) than between fossil samples in different regions (e.g., early African moderns and living Europeans). At first glance, this prediction seems reasonable given that under multiregional evolution, isolation by distance would result in the greatest amount of ancestry in any generation deriving from the same population a generation earlier. After all, in any given generation it seems most likely that any given individual would choose a mate from within the same population. Although this assumption holds true in many cases for very short periods of time, quite different results obtain for the longer spans of time encompassed even when considering “recent” human evolution. Even a small trickle of gene flow adds up very quickly.

A simple example illustrates the distinction between per-generation gene flow and accumulated ancestry. For this purpose I will use the island model of population structure, the simplest model of gene flow. Assume an “island” with an initial allele frequency of $p_0$ that receives gene flow from a “continent” that has an allele frequency of $\bar{p}$ that remains constant over time, and that this gene flow
occurs at a rate of $m$ per generation. In this simple case, gene flow is in only one direction, from the continent to the island. Each generation, the island receives a fraction $m$ of its genes from the continent and a fraction $1 - m$ of its genes from itself. After $t$ generations, the allele frequency of the island is $p_t = \bar{p} + (1 - m)^t(p_0 - \bar{p})$ (Hartl and Clark 1997). The model shows clearly that given sufficient time, the allele frequency in the island will approach the allele frequency of the continent, a simple case of genetic replacement because as $t$ becomes larger, $p_t$ becomes more similar to $\bar{p}$.

The rate of gene flow ($m$) is for each generation. The other relevant quantity is the amount of accumulated ancestry, $M$, that has occurred in total over the course of $t$ generations. For this quantity, the allele frequency in the island after $t$ generations would be $p_t = (1 - M)p_0 + M\bar{p}$ (Roberts and Hiorns 1962). Combining this equation with the one above gives the relationship $M = 1 - (1 - m)^t$. The main point here is that even given relatively low rates of per-generation gene flow could add up very quickly over time. For example, if the rate of gene flow from the “continent” were $m = 0.01$ for $t = 100$ generations, the total amount of ancestry from the “continent” in the “island” would be $M = 0.63$. For a rate of $m = 0.1$ and $t = 100$, then the total amount of accumulated ancestry would be $M = 0.99997$. It is easy to see how, under the right conditions, that genetic replacement could occur very rapidly and alter the expected pattern of biological distances over time.

### 20.5.2 Gene flow and biological distance

Although the island model is rather simplistic, the general insights obtained with it apply even with more complex migration models (Relethford 1999). The main point here is that predictions of biological distance between fossil samples based on per-generation rates of genetic exchange might be incorrect. In terms of modern human origins, several studies have examined patterns of biological distance across time and space based on cranial measures (Stringer 1993; Waddle 1994; Sokal et al. 1997). The critical biological distances are those between fossil samples over time. For example, under the African Replacement Model, one would predict greater similarity between European moderns and early African moderns than between European moderns and European Neandertals. However, the prediction that any multiregional model would show greater similarity between European moderns and European Neandertals would be incorrect.

As shown elsewhere (Relethford 1999), rates of gene flow and population size affect the predicted pattern of biological distances. Specifically, the population that contributes the most accumulated ancestry over time will determine the
pattern of relationships. As genetic evidence suggests that most (but perhaps not all) of our ancestry comes out of Africa, this means that the greatest similarity over time will be between recent modern humans and early African moderns. Thus, under either the African Replacement Model or the Primary African Origin Model, we expect to see recent Europeans look more like earlier Africans, recent Asians look more like earlier Africans, and so forth. In other words, both models generate the same pattern of expected overall biological distances. Further, the same predictions would be generated for different subsets of the Primary African Origin Model, such as a long-term pattern of interregional gene flow with a larger African population size (Relethford 1999), a pattern of dispersals with admixture out of Africa, or even a combination of these (Templeton 2002).

20.5.3 Regional continuity

Regional continuity, the persistence of traits in a region over time, is central to multiregional evolution and is the primary fossil evidence used to argue against a replacement model. At first glance, however, the previous discussions of the long-term cumulative effects of gene flow out of Africa might seem to contradict the predictions of regional continuity. If, after all, gene flow can alter our expectations of regional similarity in biological distances, then how can we expect any evidence of regional continuity? Should we expect to see gene flow erase regional continuity? No, because regional continuity of neutral traits can be caused by the interaction of gene flow and genetic drift and the fact that gene flow affects all traits, whereas genetic drift can have a different expectation for each trait (Relethford 1999).

Under a Primary African Origin, the gene flow out of Africa would affect all traits equally, and any biological distance measure that is an average of all available traits will show the closer relationship to earlier African samples discussed earlier. Because of the random nature of genetic drift, some traits will retain regional continuity in the face of the overall impact of gene flow. For this reason, we should not expect to see regional continuity for all traits, but instead only a small number of them. The balance between gene flow and genetic drift, whereby drift can oppose the homogenizing nature of gene flow, has long been known, dating back to the classic equations of Wright (1931) for the simple case of the island model. The effect of drift is the same in more complex situations, as shown by computer simulation (Relethford 2001c). These results suggest that we need to examine more closely observed patterns of regional continuity. For example, Lahr (1994, 1996) conducted a comprehensive analysis of regional continuity and found that the majority of traits analyzed did not show
regional continuity, and those that did were not found in any one region exclusively. Although Lahr argued that such findings contradict the claims of regional continuity in recent human evolution, it is clear that these patterns are consistent with regional continuity due to a balance between interregional gene flow and genetic drift (Relethford 1999, 2001c). On the other hand, it seems less likely that patterns of regional continuity would arise from under a replacement model. It thus appears that the dominant signal from the fossil record should not be regional continuity under any origin model, or in average biological distances, but instead show up in lesser degree under any model incorporating admixture. There should be no continuity under a replacement model. However, further work needs to be done in this area to determine if a replacement model with retention of ancestral traits combined with genetic drift could mimic regional continuity. We also need to address the additional complication of sampling error (i.e., we might be missing evidence of a regional trait in other regions because of small sample size).

20.5.4 Hybridization of Neandertals and modern humans

Both the African Replacement Model and the Primary African Origin Model agree that anatomically modern humans first arising in Africa expanded across Europe during the Late Pleistocene. The difference between these two models is that the latter proposes admixture between modern and Neandertal populations. If so, what pattern should we see in terms of Neandertal–modern hybrids? To date, the evidence for such hybridization has been scarce and controversial. The Lagar Velho child has been suggested to be one such example (Duarte et al. 1999; Trinkaus and Zilhão 2002), although others suggest it is simply a chunky modern human child (Tattersall and Schwartz 1999). Another possible example is an early modern human mandible from Romania that shows some archaic and possibly Neandertal features (Trinkaus et al. 2003).

If admixture occurred between Neandertals and moderns in Europe then should we expect to see more evidence of hybridization? This question provides another example of how the application of population genetics models to the fossil record might be useful in determining expected patterns. Although the expectation of hybrids might be rather high under some models of gene flow, more complex models could reveal a different expected pattern. An example in this case is the diffusion wave model (Eswaran 2002), which proposes that modern human populations expanded out of Africa and that this diffusion was due in part to a selective advantage of a modern genotype. The details of this
model involve the movement of a selective wave front where archaic genes are assimilated. For the purpose of illustration here, it is worth noting that Eswaran specifically addressed the issue of apparent hybridization in the fossil record. He noted that hybridization would be apparent only on the wave front and if the process were quick in evolutionary time, there would be little evidence of hybridization in the fossil record. For example, his computations suggest that if modern humans diffused across 3,000 km in 20,000 years, then only 10% of the fossils from that time would be expected to show clear signs of hybridization. If true, we have another example of how the fossil record might not easily discriminate between replacement and admixture models, a situation similar to that with much of the genetic evidence. Whether or not Eswaran’s model holds up under further study remains to be seen. I do not offer it here as the final word, but rather as an example of the kind of input population genetics can have in future paleoanthropological research.

20.6 Some closing thoughts

Molecular and population genetics has made an impact on the study of paleoanthropology and will continue to do so. While early work in molecular anthropology focused on the divergence of African ape and human lineages, the issue of modern human origins has dominated much of the research in the past two decades. Genetic data from contemporary populations continue to accumulate at an increasing pace, and it is likely that the growing abundance of genetic markers will soon allow questions to be answered with greater precision (Wall 2000; Marth et al. 2003). Likewise, additional genetic data on ancient populations could increase our ability to test various hypotheses, such as the maximum amount of possible Neandertal ancestry given a lack of Neandertal-like sequences in living humans (Serre et al. 2004). However, more needs to be done in developing new models and analytic methods that can potentially resolve current problems such as the recent work of Eswaran (2002) and Templeton (2002).

We also need to increase the contribution of population genetics models to the fossil record. Models need to be developed that will provide specific predictions regarding expectations from the fossil record. As noted earlier and elsewhere (Relethford 1999, 2001c), initial predictions often turn out to be incorrect or more complex than once thought. In addition, model building should not be an exercise completed in isolation from the fossil and archaeological records; models must be developed that fully utilize available information on ancient population size, migration patterns, and other demographic data. Perhaps the greatest lesson to be learned from a consideration of the fossil, genetic,
and archeological evidence is that all of these approaches are vital and necessary to truly understand human evolution. I see little point in arguing about which approach is “better” or in ignoring the potential contributions all types of data bring to paleoanthropology.

References


Wright S (1931) Evolution in Mendelian populations. Genetics 16: 97–159
21 Ancient DNA

Susanne Hummel

Abstract
Ancient DNA research, defined as the retrieval and analysis of DNA sequences from various degraded biological source materials, has promoted many biological and medical research fields during the last two decades. In particular, historical anthropology and paleoanthropology stand to benefit from direct access to backdating genetic data, as has already been shown through applications ranging from individual identification, reconstruction of kinship and marriage patterns to human phylogeny. The DNA-based prerequisites and basic methodological strategies for access to the various types of information are explained, as well as the characteristics of ancient DNA that limit the different approaches. Major restrictions arise from the degradation of ancient DNA down to fragment sizes of at the most only a few hundred base pairs. This fact links ancient DNA analysis almost exclusively to the PCR technique that enables us to deduce genetic information from degraded nucleic acids. Furthermore, ancient DNA extracts regularly consist of only a few intact target sequences, which may additionally reveal sequence deviations due to the degradation process. Both these factors make the analysis vulnerable to the generation of nonauthentic results. These pitfalls of ancient DNA analysis are explained and discussed in detail with reference to the most recent relevant literature. Wherever possible and available, suggestions for strategies to overcome commonly experienced obstacles in ancient DNA analysis are highlighted and evaluated.

21.1 Investigating the past

Today, the analysis of ancient degraded DNA extracted from forensic evidence samples and archeological specimens, hundreds and thousands of years old, is a common practice. It was the coincidence of two events, two decades ago that enabled this remarkable and comparatively rapid development. On the one hand, there were the first reports on the retrieval of ancient DNA from a specimen of an extinct quagga more than 100 years old (Higuchi et al. 1984) and a specimen of an Egyptian mummy 2,400 years old (Pääbo 1984) that electrified many scholars working on historic and prehistoric biological sample materials. On the other hand, PCR was
invented enabling the enzymatic amplification of short specific DNA sequences (Saiki et al. 1985), which proved to be a true breakthrough to a new level of information for any biological and medical discipline. Only through the PCR technique did ancient DNA research, defined as the retrieval and analyses of degraded DNA sequences from forensic evidence, museum specimens, archeological finds, fossil remains, and any other degraded trace of DNA, become viable.

In the context of population genetics, it is in general possible to deduce former states from present day genetic patterns to former states with the help of model calculations (Barbujani and Bertorelle 2001; Cann 2001). However, these approaches may suffer from heuristic assumptions that fail to prove their applicability. Biases may be caused through, e.g., unknown bottleneck situations, unknown selective forces, or the nonreliability of the molecular clock model. Therefore, direct access to historic and prehistoric genetic patterns has been a desideratum ever since ancient mitochondrial and nuclear DNA (nDNA) proved to be approachable through PCR (Hagelberg et al. 1989; Hummel and Herrmann 1991; Jeffreys et al. 1992; Gill et al. 1994).

In the early days of ancient DNA research, the connecting link for scholars from different scientific backgrounds was the fact that the investigations deal with a demanding sample material (Herrmann and Hummel 1993). However, within a short time, the field has diversified and thousands of manuscripts that may be claimed to be ancient DNA research work have been published. Among the various scientific contexts are, for example, epidemiology and public health (for reviews see, e.g., Zink et al. 2002; Greenblatt et al. 2003; Drancourt and Raoult 2005), nutritional sciences and food technology (Miraglia et al. 2004; Kato et al. 2005; Teletchea et al. 2005), histopathology and laboratory medicine (Leiva et al. 2003; Mariappan et al. 2005; Paik et al. 2005), forensic sciences (Iwamura et al. 2004; Valenstein and Sirota 2004; Budowle et al. 2005; Carracedo and Sanchez-Diz 2005; Sipoli Marques et al. 2005; Tamaki and Jeffreys 2005), human evolution (Cavalli-Sforza and Feldman 2003; Kaessmann and Pääbo 2004), paleobotany (Gugerli et al. 2005) and historical anthropology (Keyser-Tracqui et al. 2003; Hummel 2003a). Although the development of scientific questions is increasingly divergent, there is still a strong connecting link in the discussion, the question of authenticity of ancient DNA results.

Proving and disproving the authenticity of one’s own and other authors’ results is the main topic of many of the recent ancient DNA publications. However, the discussion does not include the question of whether particular ancient DNA investigations make sense or are rather an indication of a common trend to involve molecular methods as, for example, one may judge for some of
the Neanderthal investigations (Krings et al. 1997, 2000; Ovchinnikov et al. 2000; Beauval et al. 2005). It is no longer just a claim for paying attention to a list of criteria (Cooper and Poinar 2000) which—according to the authors—have to be obeyed in ancient DNA analysis. Instead, it is most welcome to read that “... the criteria were intended to assist in determining the authenticity of a study, but they cannot replace a crucial consideration of the problem” (Gilbert et al. 2005).

When we take a closer look at the criteria of Cooper and Poinar (2000), it is obvious that they reflect the typical difficulties in the analysis of ancient mitochondrial DNA (mtDNA) in particular from individual or few specimens. However, this type of investigation with its characteristic problems is far from being representative of all ancient DNA based investigations. Nonetheless, the criteria have been regarded as the ultimate solution for any type of ancient DNA investigation for quite some time, although indeed they do not apply, for example, to ancient short tandem repeat (STR) typing or the analysis of large series of samples (Hummel et al. 2000; Hummel 2003a, b).

Finally, at least some of the most recent publications seems to allow reason to prevail and raise differentiated questions, such as how to identify, how to deal with, and ideally, how to overcome results which are due to either contamination or degradation artifacts in ancient mtDNA analysis. Although not literally expressed, this new and positive trend in the authenticity debate indicates that frequently published standard terms corresponding to “... amplification products were absent in the negative controls (data not shown) ...” and “... DNA degradation pattern can be excluded to be responsible for the results ...” are now discovered at best to be wishful thinking at least as far as ancient mtDNA analysis is concerned (Hummel 2003a; Gilbert et al. 2005). However, it is more than irritating to read in the same paper of Gilbert et al. (2005)

▶ ... we encourage the extension of these questions a little further as to ask: ‘Even if I believe that the data might be slightly inaccurate, does this alter the final conclusion of the research?’ For example, although the first published Neanderthal sequences might contain a few damage-induced errors, these would not influence the conclusion that Neanderthal mtDNA is likely to fall outside the clade of modern human mtDNA. By contrast, however, such errors might affect the estimates of Neanderthal coalescence ages by a few ten thousand years and, thus, studies that intend to use the data to this end should take this into account.

With the background that the criteria of Cooper and Poinar (2000) were argued to be the sole basis for valid ancient DNA research (Hofreiter et al. 2001; Willerslev and Cooper 2005), the only possible conclusion is that the criteria were
indeed not solely an expression for the wish of good practice in natural sciences but rather served for science politics strategies to support a self-referential and considerably aggressive peer review system. Further, it is not acceptable to pass down the responsibility for validity of data to those who take them as a basis for scientific deduction (Currat and Excoffier 2004). Nonetheless, the current development which questions the omni-applicability of the list of criteria but favors the adaptation of the experimental design to the actual question (Hummel 2003a; Bandelt 2005) also gives reason to hope for a common understanding that authenticity of ancient DNA results is not ensured through the claim of excellence but through sophisticated experimental approaches.

### 21.2 Ancient DNA sources and characteristics

Genetic information may be preserved in any type of biomaterial which contained cells. In the average animal cell, basically the entire genetic information characterizing an individual is coded in the diploid set of chromosomes and hundreds to thousands of mitochondrial genomes. However, cell organelles may suffer from regular processes of partial or complete degradation already in the living organism, e.g., the decline of the nucleus in keratinous cells of hair and nails during growth. In plants, additionally the chloroplasts contribute to the genomic information. While in plants in particular, seeds are already foreseen by nature to preserve genetic information through time, a similar situation is not present in animals. However, skeletal materials, such as bone and teeth, harbor cells in a more or less intact state of preservation through the ages. Within those tissues that are characterized by a high content of inorganic material and high density, different cell types are DNA sources: osteoblasts, osteoclasts, and osteocytes, which are responsible for bone remodeling and homeostasis, as well as all types of blood cells found in the Haversian canals. Besides the cells representing the genotype of the organism whose skeletal remains were found, additionally cells of pathogens entering the blood stream may be preserved.

Typically, the majority of ancient DNA is degraded down to fragments smaller than 200 base pairs (bp) of length. This can reproducibly be shown through multiplex approaches (Figure 21.1) which simultaneously amplify DNA fragments of different lengths ranging from 100 to 350 bp.

The typical signal patterns observable in electropherograms of multiplex amplifications indicate comparatively high yields of DNA fragments smaller than 200 bp and lesser yields of larger fragments. These patterns are not dependent on possibly varying amplification efficiencies of the different primers of
preferential amplification of shorter fragments but entirely reflect the amount of target sequences. This should be considered when choosing and optimizing PCR parameters.

21.2.1 DNA degradation

Whether the genetic information of an organism will be preserved just for days and weeks or for hundreds and thousands of years depends on many chemical and biological factors that are connected in a highly complex manner and are, therefore, hardly understood and almost impossible to predict. From long-term empiric findings of many researchers and a small number of systematic studies
(Burger et al. 1999), just a few facts favoring the preservation of partly intact DNA seem undisputedly clear: low temperatures and absence of humidity, accompanied by a neutral or slightly basic pH value.

Immediately following death, these conditions are best realized in distal parts of the body, preferably in bones and teeth due to their high mineral content. When the environment of the decomposing body provides an optimal situation, the enzymatically driven autolytic decay is stopped as soon as possible; at the same time the subsequent decomposition of remaining tissues through microfauna, bacteria, and fungi has a reduced chance to effectively take place. The result is a mummification of the cells which, once it has occurred due to optimal environmental conditions, has a good chance to last for thousands of years. The DNA within the mummified cells is now protected from the initial main destructive processes: hydrolysis and the activity of exonucleases and endonucleases. However, DNA will still suffer from destruction (e.g., oxidative damage) although the order of magnitude is less effective than in the course of early decay.

As a matter of fact, the mere preservation of macromorphologically intact skeletal material indicates pH values also favoring DNA preservation. If the pH values had been as acidic as was claimed in a paper that is still often cited (Lindahl 1993) in the context of DNA degradation, the bone mineral would have been turned into the water soluble brushite and thus would no longer have existed.

Further, a more or less intact micromorphology that can be assessed by histological techniques indicates the absence of microorganisms, which is also favorable for DNA preservation.

Finally, most recent studies by the present authors indicate that the amount of preserved DNA seems to be favored not only by the density of the bony material (compact versus spongy bone) but also at the least as much by a distal anatomical region, i.e., far from the torso with its largest amounts of soft tissues. If these findings prove to be reproducible, it would not only be an enormous help in the decision to sample a skeleton but would also indicate that the very initial autolytic processes of body decay are possibly greatly underestimated concerning DNA degradation and long-term DNA preservation.

### 21.3 Information from the genome

All genetic information in a cell is called the genome, independent of its informative value, what type of cell organelle it originates from and its mode of inheritance. In animal cells, the genome regularly consists of two types of DNA
which differ from each other in many respects: chromosomal DNA, also known as nuclear DNA, and mitochondrial DNA. While nDNA is organized in the form of densely packed chromosomes that are located within the single nucleus of a cell, mtDNA is organized in up to 10 identical plasmid-like rings within each mitochondrion of a cell. This is different from the single nucleus in that there are many mitochondria per cell; depending on the intensity of the metabolic turnover of the specific tissue, it may be up to thousands.

Both types of DNA, mitochondrial and nuclear, consist of coding and noncoding regions. The coding regions, also called genes, determine protein synthesis, i.e., they are expressed through the phenotype of an individual, including pattern which can regularly not be seen, such as immunological characteristics. However, up to the present, the noncoding regions of mtDNA and nDNA have been much more the focus of ancient DNA research. This is due to the fact that noncoding regions allow a high degree of sequence and length polymorphism, which are the basis for identifying and reconstructing kinship from the phylogenetic to the genealogical level (Budowle et al. 2003; Pakendorf and Stoneking 2005; Rowold and Herrera 2005).

### 21.3.1 What does mitochondrial DNA tell us?

MtDNA is most suitable for estimating the time of divergence of two or more populations and, for example, for reconstructing migration patterns. This is due to its nonrecombinant mode of inheritance, which implies that all differences found in the mitochondrial sequence are the result of mutation events that are thought to occur in a constant manner (molecular clock). Since the sequence divergence of mtDNA represents a depth of time, it can also be easily understood why it is in principle possible to investigate present day populations in order to reconstruct events deep back in time. Depending on the choice of the sequence that is analyzed, it may allow us to discriminate individuals on the species or subspecies level or to assign the maternal lineage to a certain population.

Factors that are thought to perhaps affect the precision of the molecular clock are changes in the amount of natural irradiation (due to volcanic eruptions) and the question of the number and likelihood of back mutations, particularly at so-called mutational hot spots of the mitochondrial sequence. Additionally, in the analysis of ancient mtDNA, the indications for sequence-specific pseudo-mutations due to DNA degradation, nuclear insertions, and the particular high background of contamination through mtDNA and related questions that throw the authentic nature of the analysis result into question must be considered.
21.3.1.1 Inheritance of the mitochondrial genome

The mitochondrial genome is inherited maternally, lacking recombination and therefore representing a so-called haplotype. This is due to the fact that each female oocyte possesses a full set of mitochondria, while the sperm needs just a few to generate energy for movement. For quite some time, it was thought that mitochondria of the sperm did not enter the oocyte at all, but there is evidence that male mitochondria do enter but are identified as foreign and destroyed by the oocyte. If this process is not effective, the result is a so-called heteroplasmy, i.e., two different sequences with its particular sequence deviations are represented in an individual’s mtDNA. Heteroplasmy may also derive from replication errors during cell duplication, a phenomenon particularly often observed and well known from particular C-stretches of the hypervariable regions (Malik et al. 2002). In ancient DNA analysis, it may be challenging to distinguish between a true heteroplasmy and an artifact signal that may occur due to various reasons.

21.3.1.2 Sequence polymorphisms

MtDNA reveals a comparatively high density of genes. In humans, as many as 37 genes are represented within the total length of 16,569 bp of a single mtDNA genome. Most of these genes are involved in the metabolic turnover of the cell. Due to the nature of a gene, which codes for the synthesis of proteins, only very limited sequence deviations are possible if a nondisturbed function is to be reached. However, there are two regions within the human and animal mitochondrial genome, so-called d-loops, which are noncoding, i.e., not involved in protein synthesis. One of the d-loops, almost spanning about 1,000 bp at the nomenclatoric origin of the mtDNA, is also known as the hypervariable region (HVR) revealing comparatively extensive sequence polymorphism. Besides base exchanges, there may also be base deletions and base insertions. The sequences of two randomly chosen individuals from a population differ on average at eight nucleotide positions within the HVR. The experimental design in ancient DNA analysis from prehistoric specimens must consider possible further deviations that may lead to mismatches in the primer binding sites causing amplification failure.

21.3.1.3 Cambridge reference sequence

In order to describe the deviations unambiguously, the international scientific community has agreed to refer to a certain reference sequence. This sequence is
either known as the Cambridge reference sequence (CRS) or the Anderson reference sequence. This represents the entire sequence of a human mitochondrial genome, the first that was ever analyzed entirely (Anderson et al. 1981). However, the originally published sequence obviously reveals some rare polymorphism and sequencing errors (Andrews et al. 1999). The sequence belongs to an individual of haplogroup H, which is the major haplogroup in individuals of European descent. Since in phylogenetic terms this of course does mean that it is a comparatively young sequence pattern, there are fairly many nucleotide positions in the human mtDNA that are different from the CRS in all other haplogroups found worldwide. Moreover, the CRS reveals two comparatively rare deletions in C-stretch regions of the HVR II, which means that the length of the average human mitochondrial genome is 16,570 or 16,571 bp, respectively.

21.3.1.4 Haplotypes, maternal lineages, and haplogroups

The actual base sequence in the HVR of an individual is named as haplotype and represents the pattern specific to a maternal lineage. All individuals sharing this haplotype are members of the same maternal lineage, i.e., they are directly related in a genealogical sense, which is most likely when the haplotype is rare or even unique. Alternatively, they are at least closely related in a population genetic sense, e.g., when many people who are not known to belong to the same genealogical family share this haplotype.

The sequence differences found worldwide in the highly polymorphic HVRs of the mitochondrial genome are divided into more than 20 haplogroups consisting of further subgroups. Those groups are clusters of sequences revealing high intragroup similarities. In general, the assignment of a sequence to a certain haplogroup depends on actual bases that are present at certain key nucleotide positions. While more than 95% of individuals of European descent can be ascribed to one of seven major haplogroups (Richards et al. 1998), a similar percentage of Native Americans are ascribed to four haplogroups (Torroni et al. 1993). Particularly within far branching out ancient haplogroups, a further division into subgroups is common. Again the subgroups are defined through an additional base pattern at further nucleotide positions that is shared by all representatives of the subgroup. Since ancient DNA is suspected to promote the generation of pseudo-polymorphisms, preferably following a particular base sequence pattern due to degradation artifacts, the interpretation and deductions with respect to phylogeny and migration from the occurrence of rare or novel haplotypes must be done with care.
21.3.1.5 Phylogeny and migration

The age and spread of haplogroups enable us to draw conclusions to questions of phylogeny and migration pattern. The age of a haplogroup is basically deduced from its heterogeneity, i.e., the degree of sequence variation within the group. Basically, haplogroups that are more widespread and reveal more different branches and subgroups are older than those that are comparatively homogeneous (Figure 21.2).

In this manner, it is also possible to determine the regional origin and migration patterns of a haplogroup, which is represented over a geographically widespread area. At the place or region where representatives of the haplogroup originate from, the sequence diversity is expected to be higher than in any other region where the haplogroup is found (Watson et al. 1997).

21.3.2 What does chromosomal DNA tell us?

The analysis of chromosomal DNA sequences allows access to the unique genetic pattern of an individual. This pattern is represented through the specific

![Figure 21.2](image-url)
combination of single polymorphic genetic traits, which characterize the individual on a level that enables identification. The number of genetic traits that must be observed to enable identification depends on the degree of polymorphism that is recorded for the observed traits. Most common for identification purposes that are the basis for, e.g., the reconstruction of genealogical kinship, are so-called STRs. Those markers reveal a comparatively high degree of polymorphism, i.e., on average 5–15 alleles of varying lengths are present in a population for each STR. These markers are also suitable for population genetic purposes, although data of many individuals are necessary in order to enable conclusions with respect to phylogeny and migration.

Another target are single nucleotide polymorphisms (SNPs), which are usually biallelic markers revealing a sequence polymorphism at a certain nucleotide position. Unlike STRs, they are not investigated to reconstruct kinship, although this would be possible in principle, but in order to, e.g., determine the immunological properties of an individual. Since immunological properties of a population are subject of strong selective forces through the presence of pathogens influencing morbidity and mortality rates, regional and/or diachronic changes of allelic frequencies of SNPs allow us to draw conclusions to questions such as the spread of epidemic or pandemic events and living conditions.

In ancient DNA analysis, the accessibility of chromosomal markers is restricted through the comparatively low numbers of originally present genomes in a given sample volume. However, optimizations of DNA extraction protocols may indeed compensate for this disadvantage. Once successfully extracted, the analysis of nuclear markers is much less prone to erroneous results due to contamination given a careful experimental design.

21.3.2.1 Inheritance of the chromosomal genome

The chromosomal genome of an individual represents a novel and unique recombination of its parental organisms. Both parents contribute a so-called haploid set of chromosomes which are randomly composed from their own diploid chromosomal sets. Therefore, each locus under investigation reveals a maternal and a paternal allele. As soon as it comes to the analysis of coding regions, the genotypic level may often be determined without difficulties, but this does not necessarily hold true for the phenotypic level. Gene expression and suppression must be known in order to allow conclusions to be drawn from the analysis of the genotype of an individual to its phenotype. In ancient DNA analysis, a particular situation that is suspected to be involved in
the context of many diseases remains unsolvable: the phenomenon of compound heterozygocity. Compound heterozygocity describes the situation that two (or more) mutations, distant from each other on the same chromosome and responsible for a certain morbidity risk, have to be present in a heterozygous state, but one mutation must be located on the maternal, the other on the paternal allele. In ancient DNA with its highly fragmented target sequences, compound heterozygocity cannot be discriminated against a heterozygous state where both markers are mutated either on the maternal or the paternal allele.

### 21.3.2.2 Length polymorphisms and sequence polymorphisms

Chromosomal DNA consists of coding and noncoding sequences. From the human genome sequencing project it is known that the majority of DNA is noncoding, sometimes called junk DNA. Within the noncoding DNA about 20–30% of the DNA consists of tandem repeated sequences. Some of them, STRs, are particularly interesting since a set of them enables the genetic identification of an individual. The alleles reveal length polymorphism which is a function of the number of repetitions of the core unit. In fact the short repeats are often identical (e.g., CA$_n$ or AGAT$_n$), but the unique nature of the neighboring sequences enables us to specifically target a certain STR locus. Other than sequence polymorphisms, the variations in fragment length are not suspected to suffer from specific DNA degradation leading to a result that, although reproducible, is still erroneous. However, ancient DNA analysis must be carried out carefully since STR amplifications also reveal typical so-called stutter artifacts. Since stutter bands occur stochastically, the artifact can be overcome through multiple analyses. Further, the occurrence of this typical artifact can be minimized through optimization of DNA extraction and a sophisticated choice of PCR parameters. The multiple analyses also prevent erroneous homozygous results, which may occur due to so-called allelic dropout when only a small number of intact target sequences are present.

The analysis of ancient DNA SNPs which are spread throughout the genome, has basically to deal with the same pitfalls that are suspected of affecting the analysis of mitochondrial sequence polymorphisms. Although the risk of nonauthentic results due to contamination is much lower than in the analysis of mtDNA and, moreover, can efficiently be monitored through a suitable experimental design, the risk of false results due to specific degradation pattern must be considered in the same way as in mtDNA analysis.
21.3.2.3 Genetic fingerprints and identification

The specific allele combinations from analysis sets consisting of 5–15 STRs are also known as genetic fingerprints. As the name implies, the specific allele combinations have an identifying character. Genetic fingerprints are suitable for genealogical kinship reconstruction and for the identification and the assignment of skeletal elements.

In order to be suitable for genetic fingerprinting, the STRs combined in a set have to fulfill two major prerequisites: first, they must be located on different chromosomes in order to avoid haplotypes and to ensure a recombinant mode of inheritance. Second, they must not be linked to genes in order to avoid the force of possible selective pressure, which would be indicated by strong deviations from a Gaussian allele distribution. If these criteria are fulfilled, the value to which the so-called matching probability (Pm) can be determined depends on the total number of STRs investigated and on their respective allele frequencies in a given population. The Pm is defined as the likelihood that a second nonrelated individual reveals the respective allele combination just by chance. Typical Pm values for commercially available and customer-designed sets consisting of 9–15 STRs range from 10^{-10} to 10^{-24}. Of course, these impressive values apply to ideal populations only (panmictic, no selective pressure, etc.). However, the situation in real, possibly even inbred populations, can be simulated approximately when calculating the value that given parents will give birth to two children revealing the same genetic fingerprint. If 10 STRs are observed, the likelihood of this event ranges from 10^{-5} to 10^{-7}. These values indicate that individuals from an inbred population with a limited gene pool are still discriminated on a sufficient level.

The identifying properties of genetic fingerprints also indicate that STR typing is more suitable than any other strategy to identify possible contaminations from a wide spectrum of sources.

21.3.2.4 Reconstructing kinship and paternity testing

The reconstruction of kinship in a (pre-) historic skeletal series requires a different strategy, as if it were a living population. While in a living population, the individual age helps to a priori discriminate the parental generation from the potential descendants, this information is not available in a skeletal series. It is possible to find the skeletal remains of an individual who died in his early adulthood who fathered an individual who reached the senile age class. Although individuals who are related vertically at first degree share an allele for each
observed genetic marker, it is not possible to deduce who is the parent and who is the infant from the genetic markers. Therefore, infant individuals who have not reached reproductive age play a key role in the reconstruction of kinship (Figure 21.3).

If two individuals are found to share an allele for each STR system (=STR 1, STR 2, and STR 3), they may be parent and child. (a) From the STR typing and the Y-Haplotype (=Y1), it cannot yet be deduced who is the father and who is the son. (b) A morphological or histological age determination not only solves this problem but moreover enables us to predict the genotype of the mother who is still missing. (c) If a female individual is found who reveals the respective STR alleles as well as the suitable mitochondrial haplotype (=mt J), she can be assumed to be the mother. The likelihood of parentage can now be calculated based on the allele frequencies of the investigated STRs.

If two individuals—infant and adult—are found to share exactly one allele at each observed locus, a paternity (or maternity) test can be calculated as a deficiency case (i.e., one parent missing). If the allelic genotype of a second adult individual of the opposite sex suggests that this may be the missing parent, the likelihood of parentage for both assumed parents is calculated in the manner of two independent trio cases (Brenner and Morris 1990; Chakraborty and Jin 1993). This means that, e.g., at first the child–mother dyad is taken as fixed, enabling us to calculate the paternity index for the male individual, and then, in an analogous way, the maternity index is calculated.
In particular, if individuals from a larger burial site are investigated with respect to their genealogical relatedness, the initial search for possible child–parent dyads is facilitated by mitochondrial and Y-chromosomal haplotyping. This strategy allows an assignment of individuals to family lineages. Analogous to autosomal genetic fingerprinting, Y-haplotyping is carried out through STR allele amplifications at the Y-chromosome.

21.3.2.5 Immunogenetics and epidemiology: Toward phenotype

The analysis of nuclear genetic markers that are known to be linked to the immunological properties of an individual is comparatively new. Usually these properties are represented through SNPs which cause changes in the amino acid chain, therefore changing protein synthesis. Many of the markers are directly linked to the susceptibility of an individual to bacterial and viral infections and, as a consequence, to the susceptibility of certain cancerous growth. Furthermore, these markers are seen in the context of heterogeneous reactions of individuals to pharmacological treatments.

The allele frequencies of the markers exhibit considerable deviations worldwide, suggesting that selective pressure is responsible, as well as genetic drift and bottleneck situations. It is striking that particular European populations or individuals of European descent often reveal allele distributions that deviate very strongly from all other worldwide human populations. Not surprisingly, this is interpreted as a result of the particular European epidemiologic history during the past centuries and the selective forces that are linked to epidemic and pandemic infectious diseases events (Scott and Duncan 2001).

Many theories and hypotheses that claim to be able to name the particular historic epidemic event that is assumed to be responsible for uncommon allelic distributions are based on more or less profound linkage studies carried out on modern populations (Stephens et al. 1998; Rannala and Bertorelle 2001). However, some must be classified as mere speculations (Altschuler 2000). In any case, all studies lack direct proof. Therefore, the direct access to historic and prehistoric genotype will be invaluable. In particular, skeletal series that are known to be linked to epidemics or pandemics have an important role as genetic archives. Since the burial sites of respective skeletal series are often mass graves (e.g., plague victims), in particular, the possibility of ancient cross-contaminations due to the special burial conditions must be considered when setting up the experimental design.
21.4 Authenticity

The authenticity of an ancient DNA analysis result may be questioned due to two reasons: first, one must exclude the possibility that a contamination is responsible for the outcome of the analysis, and second, it must be proven that the result indeed represents the authentic genotype of the (pre-) historic individual and is not biased through DNA degradation or analyses artifacts. Depending on the scientific question, the actual sequence that is under investigation, the number of individuals investigated, and finally the origin of the samples, proving the authenticity of the ancient DNA analyses results may require entirely different strategies in the experimental design.

21.4.1 Contaminations

Since ancient samples usually consist of very few DNA targets, even very minor contaminations may cause false results. Contaminations may originate from various sources and may reveal various degrees of degradation. Therefore, there is no simple strategy for how to avoid and detect them.

Basically, one can subdivide four ways in which contaminations may enter the analysis:

- Cross-contamination between the ancient sample materials
- Contaminations of the sample material through former or recent investigators
- Contaminations in laboratory reagents and disposable material and
- Contaminations of either sample materials or reagents through amplification product carryover from earlier PCR analyses

21.4.1.1 Cross-contaminations

Cross-contamination of the sample material may in general happen at any time. However, two situations in particular are suspected to increase the risk of sample cross-contamination: burials at a nonindividual site (e.g., mass graves) and the laboratory processing of the samples (e.g., transfer of sample material through nonclean laboratory devices or handling mistakes). Cross-contaminations are particularly hard to detect, since all sequences involved would reveal a degradation pattern as expected for ancient DNA. Further, their occurrence cannot be
identified through the classical set of negative control samples, which includes no-template controls and extraction blanks. A possible cross-contamination event must therefore be recognizable in the analysis result itself, for example, through the amplification of DNA fragments, which would clearly indicate a second individual, e.g., three or four alleles instead of one or two.

21.4.1.2 Contamination through investigators

This type of contamination may in general happen at any time when a sample is handled. What the contamination patterns look like depends on the amount of contaminating cells and whether this contamination happened most recently or decades ago. It may vary from comparatively easy to identify fully intact DNA profiles to a hard to recognize type which resembles the typical cross-contamination. As in the case of cross-contaminations, negative control sample sets are not suitable for monitoring this type of contamination. Again, the analysis result itself must reveal indications for the possible contamination. Other than in the case of a classical cross-contamination, the contamination through an investigator is most likely (but not necessarily) a superficial one. Therefore, thoroughly removing the surfaces minimizes the risk of this contamination type.

21.4.1.3 Contaminations in laboratory reagents and disposables

Laboratory reagents may be already contaminated when purchased (e.g., primers, reaction mixes) or through handling by the investigators. In both cases, the degradation patterns of the contaminating DNA and the amount of contamination may vary strongly. If premixed reaction components are used, particularly bovine serum albumine (BSA) which is a component enhancing the Taq DNA polymerase may cause severe problems due to revealing residues of bovine DNA. This can be shown through cytochrome $b$ sequence amplifications that indicate the species. However, this type of contamination can be monitored through suitable sets of negative controls. Ideally, also the analysis result from the sample material itself provides information indicating possible contaminations.

The situation is completely different if laboratory disposable materials are concerned. Regularly, PCR reaction tubes turn out to be contaminated right from the production process, with human DNA prevailing (Hauswirth 1994; Schmidt et al. 1995) although DNA of bovine origin was already proven (Hummel 2003a).

The percentage of reaction tubes that suffer from contamination varies depending on the supplier, the brand, and the actual lot. The vast majority of
these contaminations are of mitochondrial origin. The reason for this lies in the production process, which very often includes autoclaving steps. If there are just a few reaction tubes containing cellular contaminations at this stage of the manufacturing process, these cells become lysated through the autoclaving temperature and the hundreds of thousands of mitochondrial genomes are distributed more or less uniformly. Typically, the number of tubes that show specific signals of mtDNA after 40 amplification cycles varies between 20% and 80% if fragments of less than about 150 bp are amplified.

The reason why the autoclaving step is still part of the manufacturing process, in particular in so-called high-performance (e.g., “PCR clean”, “DNA free”) and high-price brands, is that in this way the companies are able to guarantee to stay below a certificated level of DNA for each tube in the lot but need not invest in the cost of a truly DNA-free manufacturing process.

All kinds of treatments applied to amplification reaction tubes (e.g., UV treatment, bleaching, rinsing with and without ultrasonic treatment) are inefficient in the sense that they cannot overcome the problem. At best, more or less considerable reductions of the number of tubes revealing contaminations could be reached. Since the contaminating mtDNA in reaction tubes consists of just a few targets per tube and because it is typically just as degraded as ancient DNA, it is usually not even noticed in modern DNA applications.

For ancient mtDNA applications, which may also start from just a few intact targets, it is the most severe contamination problem since it is unavoidable and hard to detect. Typical sets of two or three negative controls are noninformative, simply due to their small number. If they stay blank, this is not necessarily representative for the tubes in which the samples were processed. If the negative controls reveal signals, however, this also cannot be assumed to apply for the tubes in which the samples were analyzed.

The only way to improve the situation in mtDNA analysis is through numerous reproductions of the analysis, ideally amplifying only fragment lengths considerably longer than 200 bp. In order not to have to turn down too many analysis results, which reveal ambiguous signal pattern and are suspected of deriving from the mixture of precontaminating tube DNA and ancient sample DNA, it is recommended to carry out checks with at least 30 negative controls each time a new lot of reaction tubes is in use.

Another way, although extremely laborious and inconvenient, is the use of glass tubes, handmade from high-temperature-resistant glass. They can be reused after being tempered for a couple of hours at 600°C at least. It may be necessary to elongate amplification cycle steps for efficient reactions carried out in glass tubes which, however, are positively contamination-free if the tempering is carried out.
21.4.1.4 Contamination through product carryover from earlier PCR

Product carryover, although a disaster if it occurs, is not a particular threat to ancient DNA analysis. This is because it is comparatively easy to detect and easy to prevent if some basic rules are strictly followed.

Potential product carryover would strike through the easy and low-cycle amplificability of a certain marker, while other markers of similar fragment length would not yet be amplifiable. Also, as long as not only a single sample is contaminated (which is extremely unlikely in the case of product carryover), a series of samples would reveal identical analysis results if, e.g., an extraction buffer was contaminated. If product carryover found its way into a PCR reagent, the entire set of negative controls would all show signals, thus reacting in exactly the same way the samples do.

A way to efficiently prevent product carryover contaminations, which is practiced in almost every ancient DNA laboratory, is the strict separation of pre- and post-PCR areas including a strict dedication of all devices such as pipettes, centrifuges, deep freezers, and any other instruments.

21.4.2 Degradation and analyses artifacts

Ancient DNA results may be nonauthentic due to degradation and analysis artifacts. Even though the authentic ancient DNA has been analyzed, the original sequential order of bases or the original fragment lengths may be biased. If one would just count the publications that report about degradation phenomena, one might believe that nonauthenticity due to degradation artifacts is a critical point concerning almost exclusively the base sequence (Gilbert et al. 2003; Binladen et al. 2006). This is most likely a bias as well, since only very few ancient DNA working groups have experience with fragment length analysis, which again is mainly linked to autosomal and Y-chromosomal STR typing (Keyser-Tracqui et al. 2003; Hummel 2003a). The main discussion of typical fragment length artifacts is therefore led in the forensic sciences context, where STR typing of degraded DNA samples is routine (Butler 2001).

20.4.2.1 Nonauthenticity of the base sequence

Base degradation is one of the major reasons why the analysis of authentic ancient DNA may nevertheless reveal erroneous results. The most numerously occurring artifacts are reported to be the transition C > T/G > A and A > G/T > C, while
others have been observed much more rarely or did not occur at all (C > G/G > C), at least in nDNA (Binladen et al. 2006). Another important aspect is the question of whether mtDNA and nDNA are affected to the same extent by this artifact. Although nDNA is thought less likely to suffer from degradation artifacts due to the better protection of the nucleic acid sequence through the histones, the investigations of Binladen et al. (2006) could not detect major differences. Due to its obvious relevance, the C > T transition is mainly discussed with respect to the cause of the artifact (cytosine may degrade to an apparent uracil, of which the complementary base is an adenine, which again would lead to the introduction of a thymine during the next elongation phase), its likelihood of occurrence and methods to overcome the artifact. One common strategy to prevent the amplification of DNA fragments consisting of apparent uracils is an enzymatic treatment of the DNA extract prior to the amplification (Haak et al. 2005). This strategy claims to destroy all DNA fragments that contain uracil-like degraded cytosines, which may indeed result in total destruction of all potential targets DNA sequences for the intended amplification. However, this does not necessarily mean that all the respective nucleotide positions consisting of a cytosine suffered from quantitative degradation. Further, transitions others than the C > T transitions are not reached through this approach.

Therefore, other strategies to overcome nonauthentic results due to DNA degradation would be preferable, if available. This could be valid strategies for DNA repair or unbalanced initial PCR employing a single primer that results in a linear amplification of one strand only. Further, in the case that the sample itself does not represent a mixture of DNA from different individuals, a combination of cloning strategies—which often suffer from too few clones being analyzed (Bower et al. 2005)—with direct sequencing could be informative.

Another way to monitor and evaluate analysis results that are suspected of being biased through degradation artifacts could be the implementation of cross-checks through amplification of further loci, which basically provide the same information or at least parts of this information. In the case of mtDNA haplotyping, this might be the analysis of SNP markers on the mtDNA genome, which are known to be linked to haplogroups (Torroni et al. 1992). In the case of nuclear SNPs, this may be additional polymorphic sites, which are also known to be linked to certain genes although possibly with a lower linkage rate.

### 20.4.2.2 Nonauthenticity of the fragment lengths

Typical fragment length artifacts are so-called stutter-bands and allelic dropout. Both are artifacts that occur in the course of STR amplifications and are well known
and extensively discussed throughout ancient DNA and forensic literature (Butler 2001; Hummel 2003a). Unlike sequence degradation, allelic dropout and stutter artifacts are not a direct degradation phenomenon but rather an indirect one, since their likelihood of occurrence depends on the number of intact targets that are submitted to amplification. Although the reduced amount of intact targets is characteristic for ancient DNA extracts, the artifacts occur as well in modern intact DNA amplification with few targets throughout the first amplification cycles. But even when very few targets are submitted to the reaction, the generation of this type of artifacts remains a stochastic event that is—other than in the case of base degradation—not due to alterations of the target but an amplification artifact. Consequently, the mere multiple repetition of the analyses enables us to evaluate the authenticity of a result on the basis of the reproducibility. Additionally, there are further strategies to positively influence the tendency of the amplification reaction to create these types of artifacts. The most obvious one is to optimize the DNA extraction in order to increase the number of targets. This in particular helps to avoid allelic dropout. In order to minimize the generation of stutter bands, the relaxation of the elongation temperature has proved to be successful. Since the Taq DNA polymerase activity is negatively influenced through the measurement, the time given for elongation must possibly be prolonged.

In any case, the fact that STR amplifications are regularly carried out in multiplex approaches means that cross-checks are built into the analysis, since the evaluation of authenticity is clearly simplified through the patterns of the entire set of loci amplified (Figure 21.4).

### 21.5 Basic strategies in experimental design

There is no general strategy for how to carry out an ancient DNA analysis, but the most determining factors for the experimental set up are the sample material itself, the scientific question, and the properties of the DNA sequence in focus. These factors should be considered in making the decision for possibly necessary preexperiments (e.g., check for contamination rates in amplification tubes), the choice of appropriate sets of negative and positive controls, and the design of the amplification and analysis strategy.

The final experimental strategy ideally combines two aims: the greatest likelihood to achieve a result that at the same time can positively be proven to be authentic (Figure 21.4).

Since authenticating results are achieved through entirely different strategies depending on whether for example human or animal DNA is under investigation, or, whether a single sample or series of samples are analyzed, whether
mitochondrial or nDNA is the target, or whether base sequences of fragment lengths carry the information, it is obviously unnecessary to discuss the criteria for ancient DNA analysis (Cooper and Poinar 2000), since they do not consider any of the differences mentioned and are therefore overemphasized. An experimental design following these published regulations would most of all match the analysis of human mitochondrial HVR DNA from a single sample. But even for this particular application, following all criteria and regulations does not result in...
a full proof of authenticity since cross-contaminations and contaminations through earlier investigators cannot be ruled out.

21.5.1 Sample preparation

The sample preparation has to be devoted to two main targets:

- To remove possible contaminations and
- To prepare the sample as optimally as possible for DNA extraction and further amplification reaction.

The first aim may only be achieved in the case of superficial contaminations, i.e., most commonly, contaminations of present or former investigators. If the surface of the sample is removed profoundly, the risk of this type of contamination is minimized. How the second aim is achieved depends on the sample material. However, in general, any increase in surfaces through, e.g., crushing the sample prior to a chemical treatment has proven to improve later analysis results. Further, depending on the actually chosen extraction technique, also parameters, such as centrifugation forces and the relation of sample material to incubation buffers and reagents, are crucial. It is recommended to optimize a standard protocol with respect to given sample properties before valuable sample material is processed.

21.5.2 Optimizing DNA extraction

Optimal DNA extraction protocols are obviously a highly individual matter depending on the exact biochemical state and composition of the sample material. However, standard protocols are valuable to start from. Basic aims for standard protocols are

- To maximize the amount of DNA
- To minimize DNA degradation in the course of the extraction procedure and
- To minimize inhibiting substance residual in the DNA extract

At present, these goals seem at best to be achieved through an automated DNA extraction using magnetic beads, although the amount of DNA in the extracts stays somewhat behind, e.g., phenol–chloroform procedures. However, the advantage of the DNA extracts being practically free of inhibitors, as can be demonstrated through real-time PCR, outbalances a possible loss of DNA. In
general, the use of any type of automation at this stage of sample processing is advantageous since the handling of samples is minimized and thus the risk of introducing contaminations.

21.5.3 Amplification strategies

In general, there is no optimal amplification strategy, but the basic experimental design must be deduced from the scientific question and the sequence in focus (Figure 21.4). However, there are some primary aspects which should be considered.

- The more polymorphisms are characteristic for the investigated sequence, the lesser cross-checking experiments are necessary. Ideally, the amplification result is individual in the sense of identifying since this enables an efficient check for any type of a possible contamination event. If, for example, nuclear STR typing is carried out, only a minimum of negative controls is necessary. It may be recommendable to deviate from classical amplification parameters, in particular concerning the elongation temperature. A decrease in the elongation temperature or the shift to a two-step PCR consisting of a denaturing and a prolonged annealing phase only may result in a remarkable decrease of stutter-bands, which is an STR-specific amplification artifact even in modern DNA amplifications.

- If the markers of interest are polymorphic but far from revealing results specific for an individual. In this case (e.g., any type of biallelic SNPs or short deletions), the aim should be to link the result to an identifying one. This is realized through integration of the primers for the, e.g., SNP amplification to a multiplex assay designed for STR-based genetic fingerprinting. This has proved successful for nuclear markers (Bramanti et al. 2000; Fulge 2005; Hummel et al. 2005; Puder 2005), although it may require multistage fragment-length determinations including a restriction fragment length polymorphism (RFLP) analysis. If the biallelic markers are multiplexed with nuclear STRs, only a minimum set of negative controls is necessary.

- The amplification of mtDNA, which is regularly represented in ancient DNA extracts by more intact targets, is hard to carry out in the same approach with the amplification of nDNA. One reason is the different demands for optimal amplification cycles. Moreover, the aim of mtDNA investigations is usually a sequence analysis which—due to technical reasons—cannot be
combined to STR typing. Prechecks of contamination rates for different brands of amplification tubes are highly recommendable. They should be carried out with the primers that are intended for the actual investigation using high cycle numbers (>45). In general, the amplified sequences should be as long as possible (preferably >250 bp) due to the observed decrease of intact contaminating targets. The primers should reveal sensible mismatches against species that are not intended to be amplified but are suspected to be present as contaminating targets (human, working animals). It may be necessary to launch experiments that check sequence specific degradation patterns. Although mtDNA is regularly represented by more targets than chromosomal DNA in an ancient DNA extract, the given situation concerning degradation pattern and contaminations in amplification tubes complicate the proof of authenticity remarkably. Due to these facts and the nonindividual character of the polymorphisms, even in the HVRs of mtDNA well-adapted sets of negative controls are necessary, since typical sets of negative controls do not apply.

### 21.5.4 Enhancing specificity and sensitivity of a PCR

The specificity of a PCR, i.e., the exclusive amplification of the targeted sequence, is a basic prerequisite, in particular if the following analysis includes a so-called Taq cycle sequencing reaction. However, any other type of ancient DNA analysis suffers from the generation of unspecific by-products as well. This also holds true even if no confusion with the targeted sequences is possible due to, e.g., entirely different fragment lengths. The reason for this is a decrease in sensitivity as soon as by-products including primer dimers are generated in the amplification reaction, since the resources for the reaction (Taq polymerase activity, primers, dNTPs, and buffer) may be exhausted through this competitive situation. The consequences are that the targeted product may—although present—not or only weakly reach the detection limit, since the by-reaction already causes the attainment of the amplification plateau.

Although an optimization of the annealing temperature positively influences the specificity of the amplification reaction, however, a highly specific reaction may even so not be reached or compromise the reaction efficiency if the annealing temperature approaches the melting temperature of the primers.

A particularly valid instrument to increase the specificity and therefore the sensitivity of the reaction is given through primer design.
The criteria for a good primer design are

- The primers should reveal a minimum length of 20 bp if possible, preferably 23–30 bp
- The binding energy of the 5’-end must increase the one of the 3’-end in order to enable elongation only if the entire primer matches
- 3’-end primer dimers and hairpin formation must strictly be avoided since these formations are elongated causing generation of by-products
- Should not reveal mismatches to the intended target within at least 3 bp from the 3’-end

If these demands can be fulfilled for a primer pair, the amplification products considerably increase in quality and quantity.

References

Fulge M (2005) Laktosetoleranz in der bronzezeitlichen Lichensteinhöhle—Molekulargenetischer Nachweis des Polymorphismus C/T 13910 an prähistorischer DNA. Staatsexamensarbeit, Göttingen
Hummel S (2003a) Ancient DNA typing: Methods, strategies and applications. Springer, Heidelberg
Hummel S, Bramanti B, Schultes T, Kahle M, Haffner S, Herrmann B (2000) Megaplex DNA typing can provide a strong indication of the authenticity of ancient DNA amplifications by clearly recognizing any possible type of modern contamination. Anthropol Anz 58: 15–21


Torroni A, Schurr TG, Yang CC, Szathmary EJ, Williams RC, Schanfield MS, Troup GA, Knowler WC, Lawrence DN, Weiss KM, Wallace DC (1992) Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations. Genetics 130: 153–162


22 Paleodemography of Extinct Hominin Populations

Janet Monge · Alan Mann

Abstract

This review is concerned with a critical evaluation of the paleodemographic studies of the hominin lineage prior to the development of agriculture. Because of the potential this research has for the generation of data about birth spacing, mortality, life span, sex ratio, patterns of fertility, and maturation, the study of the demography of earlier human populations has attracted much attention. The very limited and fragmentary sample sizes, however, combined with many uncertainties about depositional patterns have resulted in major difficulties in the development of generally accepted hypotheses. An additional problem involves the choice of comparative samples for these extinct hominins. Are the data from chimpanzees or modern humans more appropriate in these reconstructions?

Published reevaluations of a number of widely accepted concepts, such as the simple association of life history variables with structures like gross body or brain size, have made them increasingly untenable. Recently collected data on modern humans and free ranging chimpanzees cast some doubt on the idea that these two primates have different timing in their maturation and life span. Some life history parameters, however, such as potential life expectancy, time at maturation, and age at weaning, may very well be amenable to more complete understanding.

22.1 Introduction

Demography is the study of population structure, its size, composition, and related features such as sex, age, geographic distribution, and other environmental, social, and cultural factors. The successful application of demographic studies in modern human societies relies on the accumulation of data on all members of a population or, more commonly, on a calculated sampling of the population (Yaukey and Anderton 2001). In contrast, paleodemography has been characterized
as the field of inquiry that attempts to identify demographic parameters from past populations derived from archeological contexts (Hoppa 2002 p 9) and is primarily based on the analysis of human skeletal materials. The preservation of human skeletons in contexts of even the relatively recent prehistoric past, where agriculture and permanent settlement are present, is variable and dependent on environmental conditions such as soil type and pH, climate, and destructive geological events. Additionally, cultural factors, such as cremation, cannibalism, mass burials, indifference to the disposition of dead infants, and many other behavioral factors also dramatically influence the preservation of human bones. The present review is limited to an examination of the paleodemography of our earlier ancestors, prior to the development of agriculture.

Reconstructing population structure from this more distant past is much more problematic. Preagricultural hominins, it has been postulated, relied on a subsistence base founded on the gathering and collecting of wild foods, insects and small vertebrates, and the hunting of larger animals. Scattered and seasonally available food sources meant that for the most part, these peoples did not have permanent settlements. Rather, they established temporary camps, exploiting the local resources and moving on when these were depleted. The duration of a stay at a particular locale was dependent on the season, abundance of resources, size of the group, and other variables. The groups were small in size (Weiss 1973) often no larger than 30–50 individuals. Temporary encampments and continuous movements imply that deaths would have occurred at different places and that, except for unusual circumstances, we should not expect to find remains of a number of individuals at a specific locale. The expectation is that many deaths will occur during movements and the body left or interred, along the route. Further, deaths would have been seasonal, occurring during those times of the year when resources were scarce; this implies that hominin remains are more likely to be found at some locales and not others. These generalizations about early members of the human lineage, however, are overwhelmingly based on the anthropological study of living gatherers and hunters, and the accuracy of these observations when applied to earlier, now extinct humans, is uncertain. There is, for example, no certainty that all earlier hominins actually were gatherer/hunters; they may have been more specialized in dietary choice and habitat and thus were organized in a different but unknown fashion. As there are periods in the hominin fossil record which indicate that multiple hominin species coexisted, perhaps in geographically close proximity, it remains possible that early hominin demography was based on very different adaptive patterns. Further, prior to about 115 ka, the sample of hominin skeleton bones available for study is very small. Only after this time did the introduction of the deliberate burial of the dead result in many more bodies being preserved, and numbers of more or less
complete skeletons excavated. Finally, study of the earliest known members of the hominin lineage, mainly early members of the genus *Homo* and still earlier species of the genus *Australopithecus*, is fraught with difficulty (for the moment, fossil samples of the very earliest identified hominins, *Sahelanthropus*, *Orrorin*, and *Ardipithecus* are much too few even to be considered here). Bones of these creatures are discovered in very different circumstances than later-in-time hominins. These early remains are usually broken and very fragmentary; multiple bones from the same individual are rarely discovered. The deposits in which they have been preserved have been formed by long-term geological processes; and they are not directly associated with archeological accumulations. The reconstruction of the taphonomic circumstances that led to the hominin bones being deposited in the geological sediments in which they were found is complex and often ends with uncertain results.

The study of general paleodemographic variables is therefore limited by the recovered skeletal and fossil samples and their contexts. These do not yield valid population samples for the construction of the sort of demographic profiles possible in the study of extant humans. Thus, the techniques of analysis and the modeling of data accumulated on these skeletal samples, even under the best of circumstances, can provide only the most preliminary and shadowy details about population structure and basic life history variables.

### 22.2 Basic parameters for paleodemographic study

#### 22.2.1 Estimation of the age at death of skeletal remains

A demographic study of a skeletal sample begins with the assessment of both the age at death and sex distribution of the skeletal material. These data are absolutely essential for the construction of other population statistics. Aging and sexing of skeletal samples of living humans has been the subject of extensive research focused on the development of techniques that can yield results with a high degree of accuracy. The problems associated with aging and sexing and the ways in which the resultant data have been modeled are discussed in a recent summary by Hoppa and Vaupel (2002).

When the sample is of nonmodern skeletal materials (Neandertals or australopithecines, for example), the criteria for judging these factors are far from certain. Establishing the age at death of a skeleton relies on knowledge of the life history of the species concerned and the biological pathways within which these hominins grew and matured. There remains uncertainty and debate about the
timing of maturational events in the dentition or the skeleton and about whether models based on living humans or chimpanzees provide greater accuracy. Recent considerations of the variation in growth that characterizes chimpanzees and living humans adds a further complication to these estimations (Zihlman et al. 2004; Monge and Mann 2006).

Although they are often considered together as part of a single aspect of paleodemography, it is our contention that potential life variables, including life span (often referred to as the study of life history), should be dealt with separately from a consideration of the achievement of this life potential (which is often viewed as a central focus of paleodemographic studies). Potential life span is an evolutionary biological phenomenon (with important implications for the evolution of culture) whereas achievement of that life potential represents the complex interaction of biology and culture. Carey and Judge (2001) aptly summarized this dichotomy when they observed that modern humans probably achieve a level of life potential far earlier than this potential can be routinely achieved. Through biomedicine, diet, sanitation, insect control, and other cultural appliances, modern societies have been attempting to reconcile these two elements: to make life potential and individual life span the same phenomenon, not just for a favored few but for all humans. In reality, while all living human populations have the potential for the same potential life expectancy, most people do not achieve this. Comparisons of differences between populations are one way to analyze the contributions and adequacy of cultural mechanisms to achieve this potential and are of possible use in paleodemography. For example, using a sample of 768 dental individuals spanning the course of most of human evolution, Caspari and Lee (2004) have argued that increased cultural complexity was primarily responsible for the increases in life span during the course of the latter phases of human evolution.

When and under what circumstances members of the human lineage achieved a modern humanlike potential life expectancy is a matter of some debate. This includes debates on all phases of life history, from infancy through childhood and adolescence to adulthood. More fundamentally, even statements that the potential of life expectancy in humans is double of that in chimpanzees must now be questioned, especially the statement that potential life span in humans is double that of chimpanzees. Maximum age for humans is frequently quoted as 90–100 years and 50 years in chimps (Sacher 1975; Hawkes et al. 1998; Bogin 1999). However, our knowledge of the life spans of chimpanzees is presently limited to a few known-age captive animals and is hampered by the paucity of field observations of wild chimpanzees, even though some have been observed for 40–50 years. Cheeta, the chimpanzee who was featured in the *Tarzan* movies, is still alive (early 2005) and healthy at 71 years of age.
The chimpanzees of Mahale have been studied for 34 years and are now providing a limited demographic data set, which suggests that death among older animals occurs between 31 and 48 years (Nishida et al. 2003). This is not radically different from the profiles of average age at death in many pre-1900 human populations and for what has been estimated for most of human history (Gage 2000; but see also Ascadi and Nemeskeri 1970). The Mahale chimpanzee study also reports that approximately 25% of older females had a postreproductive life span. This phenomenon, the equivalent of postmenopausal human females, has traditionally been cited as a uniquely human life history event. From these observational data, it is clear that this postreproductive phenomenon can no longer be considered a unique event in human life history; theories that have been associated with this pattern, e.g., the grandmother hypothesis, must be reconsidered in light of this data (Hawkes et al. 1998; Packer et al. 1998; Blurton Jones et al. 1999; Alvarez 2000; Hawkes 2003; see also Mithen Volume 3 Chapter 23), although it is possible that the extent of these postreproductive years might be greater in humans (Blurton Jones et al. 2002).

Since the process of aging causes an exponential increase in virtually all pathologies (Harrison and Roderick 1997), evolutionary selection for increased longevity must target the genes responsible for aging and not focus exclusively on a reduction in fatalities associated with pathological conditions. In baboons, longevity has a reasonably strong genetic component (Martin et al. 2002) and it is possible that chimpanzees and humans have the same potential for longevity; via cultural mechanisms, humans have reduced the cumulative effect of pathological conditions directly associated with aging. Vaupel et al. (1998) have provided a more detailed discussion of the effect of increases in longevity on human demography.

The causes of death in past populations are also difficult to determine. Nishida et al. (2003) have calculated that in the Mahale chimpanzees, just under half of deaths are caused by disease. This appears to exclude general causes associated with death by senescence. In a study of a population of chimpanzees in Bissou Guinea, Sugiyama (2004b) reported that under conditions of ecological stress, in this case deforestation, specific subsets of the group were more likely to die: infants (0–3 years of age), juveniles (4–7 years), and active adolescents (8–11 years).

The primary cause of death reported for the Hadza, a gatherer/hunter group in Tanzania, between the years 1985 and 1997 (Blurton Jones et al. 2002), appears to be disease, here also excluding death associated with the state of just being old. Sugiyama (2004a) has reported on the extent of injury in a forager-horticultural group, the Shiwiar (Ecuador). The results of 678 injuries suffered by 40 individuals indicates that trauma is likely, in fact, common at all stages of life and that
without group provisioning, the effects of these injuries would range from debilitating to lethal. If the same sorts of patterns existed in the past, it is reasonable to suggest that provisioning may have played a role in the social system of earlier hominin taxa. Berger and Trinkaus (1995) reported on the incidence and anatomical position of bone fractures in Neandertals. The position and frequency indicated that the Neandertals were living a challenging life style with a preponderance of injuries on the upper part of the body. The Krapina Neandertal collection also shows a significant number of nonlethal injuries (Kricun et al. 1999), some of which appear to have demanded provisioning of some sort, for example, the Krapina 34.7 parietal in which a large cranial depression shows a considerable amount of posttrauma healing (Mann, in preparation).

Except for the widespread signs of bacterial infection on the Middle Pleistocene Kabwe specimen from Zambia, virtually nothing is known about the cause of death in fossil hominins. Almost nothing is known about the cause of death in most recent human archeological samples as well (see, for example, a description of paleopathology in Lovell 2000). Pettitt (2000) argued, using the oldest aged Neandertal skeletons, primarily Shanidar I and La Ferrassie 1, that life was difficult in the past and that death primarily resulted from repeated sustained trauma. He based this conclusion on the general robusticity of the Neandertal skeleton, including dense bones and strong muscle markings which suggested that the body was repeatedly challenged in life. In contrast, X-ray analysis of the entire fossil skeletal sample from the site of Krapina (Kricun et al. 1999) concluded that the skeletons are of healthy individuals, with the dense bone structure indicative of an active lifestyle.

On the basis of assemblages of bones showing a lack of representation of very young and old-aged individuals, Bocquet-Appel and Arsuaga (1999) have argued that at the two largest hominin fossil sites in Europe, Krapina and Atapuerca (SH), the hominin accumulations were the result of a catastrophe. Cannibalism has been considered a possible cause for the accumulation at Krapina (White 2000); that at Atapuerca (SH) has been attributed to trapping of hominins by bears (but see also Bermudez de Castro and Nicolas 1997). The age distribution at these sites, with an overabundance of adolescents and young adults, fits the paleodemographic profile of skeletal materials caught in a catastrophic event (Paine 2000). The death profile of many catastrophes, for example, the tsunami in the Indian Ocean in early 2005, often illustrates the overwhelming overabundance of children including adolescents in the death assemblage; estimates calculate that at least one-third of the deaths in South East Asia were of children.

Carnivore activity is considered a possibly significant contributor to the australopithecine fossil assemblages in the South African dolomite caves (Brain 1981). In this instance, the preponderance of young (but not infant) individuals
might be most reasonably explained by differential predation of the young, something that is documented for African ground dwelling baboons (Mann 1975). Some sort of catastrophic event may apply to the proposed simultaneous death assemblage in the A.L. 333 site at Hadar, Ethiopia.

In a series of papers, Bocquet-Appel (1982, 1985, 1986, 1996) has emphasized the difficulties of achieving dependable paleodemographic parameters from the aging of skeletons. He has argued that no matter how they are modeled, the population profiles can only incompletely represent the living group from which they derived and only in one small slice in time. Further, considering the small, nonrandom sample sizes that are the usual subject of study, if the population from which the skeletons derive are in the process of demographic change, for example, in a general or even brief trend of population increase, then the resulting death assemblage is likely to show a preponderance of young individuals relative to the number of adults. The same pattern can also be explained by a general increase in population in-migration. Thus, without large samples sizes covering a longer time period, it would appear inappropriate to propose any broad generalizations.

Because of the nature of the fossil record, the bulk of efforts at aging concentrate on the dentition, the most likely element to preserve, although other parts of skeletal anatomy have also been used to age fossil skeletal elements.

In forensic anthropology, the determination of as precise an age at death as possible is crucial for individual identification. In paleodemography, this level of accuracy is not required and the use of life table modeling is considered to be effective. Nevertheless, derived age structure must be used to model the population appropriately in the first place, with the assumption that age-at-death differences in populations form an inherently interesting question. In human evolutionary studies, the age at death distribution question is an interesting one since it allows us to attempt to understand evolutionary process from the perspective of mortality.

One of the major critiques of paleodemography discussed by Bocquet-Appel and Masset (1982, 1985, 1986, 1996) involved the assumptions made in the process of aging a skeleton. This initiated a debate evaluating the criteria employed in the estimation of the age at death of skeletal samples (Bocquet-Appel and Masset 1982, 1985, 1996; Van Gerven and Armelagos 1983; Bocquet-Appel 1986; Greene et al. 1986; Konigsberg and Frankenberg 1994). The discussions focused on the use of aging standards based on the biases already present in the reference samples to such an extent that the resulting aging profiles mimicked the age distribution of the reference sample. Bocquet-Appel and Masset’s (1982) principal concerns about the limitations of paleodemographic studies include the assumptions that the populations from which the skeletons derived were stable,
life history patterns were the same throughout human evolution, and mortality patterns in the past can be understood using recent human and primate populations. Many of these same points were made by Howell (1982) on a paleodemographic study of the Native American Libben Site, Ohio (Lovejoy et al. 1977). A recent review by Milner et al. (2000) evaluates the issues first raised by Bocquet-Appel and Masset (1982).

These discussions require critical appraisal of the published literature on the estimation of the age at death of the skeleton (for comprehensive reviews, see Katzenberg and Saunders 2000; White 2000; Hoppa and Vaupel 2002). In many cases, these standards have been developed in the USA on the Hamann-Todd Osteological Collection (curated at the Cleveland Museum of Natural History, 1 Wade Oval Drive, Cleveland, Ohio 44106) and the Terry Collection (curated at the Department of Anthropology, Smithsonian Institution, Museum of Natural History, Washington DC 20013), reference collections of known age and sex at death individuals.

Early research that presented models of the life span of fossil hominins, summarized by Ascadi and Nemeskeri (1970), examined a variety of fossil hominin materials from Europe and Asia. Later, more elaborate discussions (Mann 1968, 1975; McKinley 1971) focused on the large fossil collection of *Australopithecus robustus* from the Swartkrans site in South Africa. On the basis of observed similarities in the pattern of dental development among the significant sample of immature individuals in the Swartkrans fossil collection to that established for modern humans, Mann argued that these early hominins matured in the prolonged period then thought to be unique to *Homo sapiens*. This conclusion was criticized by Sacher (1975, 1978), whose research on the correlation between brain size and longevity suggested that the life history trajectory of the small brained robust australopithecines was more apelike than humanlike. Smith (1989) expanded on this work, arguing for a strong correlation between dental development and brain size. More recent data on primate biology has made some of Smith's conclusions untenable. For example, factoring in variations in diet, Godfrey et al. (2001) have shown that cranial capacity alone is an insufficient predictor of dental development within primate taxa. Although not frequently quoted in anthropological reviews of life history, Carey and Judge (2001), using a large longevity sample, have shown that brain size is only one of a number of central aspects of life history. Gage (1998) has also pointed out that in order to extrapolate to primate life history, more information is needed on patterns of variation, especially environmentally induced variation. Finally, in this context, Leigh (2004) has shown that in primates, brain growth is an extremely complicated phenomenon and that adult brain size alone is not well associated with dental maturation or with the length of the juvenile growth period.
Age profiles of more recent fossil hominin accumulations are not very different from those of living peoples. Take for example, the skeletal series of Natufians (the pre- or incipient agricultural groups of the Levant) analyzed by Karasik et al. (2000). In this study, using a traditional comparative series as the standard, the mean age at death was 31.5; in contrast, applying aging standards derived from local Sinai Bedouins produced a mean age at death of 36.5. Very few individuals were placed in the older adult category of 45–50 years of age. Unfortunately, the fossil record of earlier phases of human evolution, with fragmentary, incomplete bones of few individuals generally deriving from many generations cannot be subjected to this sort of critical comparative study.

Probably the best studied fossil assemblages for which paleodemographic profiles have been developed are from Krapina and Atapuerca (SH). Each site contains a large number of individuals; for Krapina, between 75 and 92 (Wolpoff 1999) and at Atapuerca, a minimum of 32 individuals (Bermudez de Castro and Nicolas 1997). As mentioned previously, both sites have been interpreted as possessing assemblages indicative of a catastrophic event. This idea for the accumulation of the sample at Krapina was supported by White (2000) and Trinkaus (1985) but criticized by Russell (1987a, b). At Atapuerca, Bermudez de Castro and Nicolas (1997) have argued against the possible catastrophic event proposed by Bocquet-Appel and Arsuaga (1999). Age estimations based on tooth development and emergence and occlusal wear, the elements usually examined in deriving age at death determinations in fossil hominins, resulted in a relatively low average age at death at both sites. Profiles from both Atapuerca and Krapina reveal an age at death distribution with high adolescent and young adult mortality and a dearth of older individuals. In an extensive examination of Neandertal paleodemography, using 206 European and Middle Eastern Neandertal remains and comparisons to 11 modern human populations including gathering and hunting peoples, agricultural peoples, and archeological collections, Trinkaus (1995) concluded that Neandertal populations were under extensive environmental stress that resulted in a unique population profile. The elements of this profile included high levels of infant mortality extending into young adulthood, with few or no individuals surviving into older age categories (+40 years). This latter feature of the Neandertal mortality profile is unique among all the modern human populations used in his analysis, including archeological samples. Trinkaus (1995) considers other explanations for this mortality pattern and attempts to minimize the effect of the Krapina and Vindija samples on the resultant profile. These two Croatian sites account for the bulk of adolescent specimens in the overall Neandertal sample, emphasizing the possibility of alternative explanations for the unusual mortality profiles represented at these sites.
If the Middle Pleistocene Atapuerca (SH) site and the Late Pleistocene analyses by Trinkaus (1995) for Neandertals are acceptable, using life tables from modern human populations it would not appear possible for these extinct hominins to replace their numbers in each subsequent generation. In another study, Trinkaus and Thompson (1987) concluded that there were no older individuals in Neandertal populations. This leaves us in the perplexing situation of having to explain why, if these death assemblages are representative of living populations, these hominins did not become extinct after just a few generations. Clearly something is wrong with these analyses. In this context, Ogilvie et al. (1989) concluded that dental enamel hypoplasia indicates that Neandertals sustained continuing stress from weaning to adulthood, resulting from an inadequacy of foraging technique. One major problem with this analysis is the “osteological paradox” (Wood et al. 1994) in which high frequencies of stress indicators on the skeleton are actually an indication of the adequacy of cultural and/or biological mechanisms in the individuals and populations at risk, to buffer the negative environmental effects.

For Middle and Upper Pleistocene hominin samples, such as those from Atapuerca (SH) and the European Neandertals, it might be possible to use life history parameters that have been described for modern humans (but see Dean et al. 1986; Bermudez de Castro and Rosas 2001; Ramirez-Rossi and Bermudez de Castro 2004). For earlier hominins, the issue is very controversial. Caspari and Lee (2004) attempted to circumvent the problems associated with life history differences in human evolution by using dental attributes to grossly divide the hominin fossil record (including members of Australopithecus) into younger and older age categories, pointing out that this division will reasonably apply to all the taxa, regardless of possible differences in growth and development.

22.2.2 Determining the sex of skeletal materials

There are also significant limitations in the identification of the sex of a skeleton. How applicable to extinct hominins are the anatomical criteria developed on modern human samples for determining the sex of individual specimens? In general, the sex determination of earlier hominins is based on the anatomy of the ox coxae and on comparisons of the level of robusticity of the preserved postcranial bones in the sample. In the case of pelvic bones, unambiguous identification is often impossible, even when the bones are relatively complete and undistorted (Rosenberg 1988). Gracilization related to evolutionary change in fossil samples over time can confound the recognition of sex differences based
on skeletal robusticity. For example, the locus H mandible from the Lower Cave, Zhoukoudian, has been identified as a female in comparison with other mandibular specimens. However, it is possible that this Locus was deposited later in time than the other fossil bearing loci, and the more gracile nature of this fossil may be indicative of a male from a later time (Mann 1981).

The clear implication of these problems is that the recognition of sexual dimorphism within an extinct hominin species can be very difficult. Plavcan and Cope (2001) have attempted to determine the validity of species in the fossil record using metric criteria and sample variation. Other researchers have also examined this problem (Lockwood et al. 1996, 2000; Rehg and Leigh 1999), and much of this literature has recently been reviewed by Scott and Lockwood (2004). More broadly, Plavcan (2001) has reviewed the pattern of sexual dimorphism in primates. These studies consider the evidence for the presence of a range of variation that can be reasonably accommodated within a valid species; within such metrically defined species, sex differences are established along the continuum. Ultimately, however, since sample sizes of earlier hominins are very small, both sexes are often collapsed into one. As a consequence of this, the determination of sex in hominin fossil paleodemography has attracted less attention than the establishment of criteria associated with aging.

However, the identification of the sex of fossil skeletal materials can yield significant data on population structure. This is because knowledge of female mortality is crucial to understanding patterns of fertility. Using comparative data from chimpanzees and modern humans groups, Lovejoy (1981) hypothesized a demographic transition at the origin of the hominin lineage, primarily based on decreasing birth intervals. Dall and Boyd (2004) have suggested that lactation in mammals probably evolved as a way to minimize the impact of fluctuations in food resources. Thus a decrease in time spent in lactation and its subsequent effect on birth interval may have been associated with a relatively more consistent food supply in human evolution. Birth interval, age at menarche, and age at first birth (first parity) are highly variable in both chimpanzees and humans (Eveleth and Tanner 1990). Sugiyama (2004b) summarized data on chimpanzees which showed that captive animals display reduced times in each of those categories associated with fertility; in contrast, fertility patterns among wild chimpanzees are quite varied.

Although sexing the skeleton is relatively easy in strongly dimorphic species, the process in modern human populations is much more difficult. Populations and species have varied degrees of dimorphism, and it can be expressed in different ways. White (2000), in a summary of the techniques used to sex the skeleton in recent archeological populations, recommends seriating the specimens and then determining the most appropriate features within that population to use

In earlier, nonmodern hominins, the presence and extent of dimorphism is difficult to understand and quantify. Frayer and Wolpoff (1985) described various models for the identification and comprehension of sexual dimorphism in human evolution. More recent research has generated new models for understanding the context of dimorphism across the hominin lineage and among vertebrates in general. For example, if late secondary sex characteristics are primarily influenced by the production of testosterone, then an understanding of reduced dimorphism in hominins might be explained by a selection model directed toward a decrease in testosterone production or changes in its target cells. Thus, the immune suppression effect of testosterone production is reduced, influencing longevity and the role of fathers in care giving might have increased as male–male competition based on testosterone levels was reduced (Wingfield et al. 1997). This is a plausible explanation for reduction of sexual dimorphism and provides an explanation for the evolutionary role of monogamy in human evolution.

In general, however, the study of the regulation and evolution of sexually dimorphic characteristics from a physiological and genetic perspective is just beginning. It is reasonable to infer that some combination of hormonal and genetic factors plays an important role. For example, Skuse (1999) proposed a mechanism of genomic imprinting of the X chromosome as a possible mechanism, and Haqq and Donahoe (1998) reviewed the literature on individuals diagnosed with sexual ambiguity to construct a more holistic model of the factors contributing to the attribution of sex. Significant differences exist in the brains of human males and females, and sex differences are present in many biological systems (Maguire et al. 1999; Goldstein et al. 2001; Allen et al. 2003; Dubb et al. 2003; Raz et al. 2004; Shah et al. 2004; Vawter et al. 2004). Bolnick and Doebeli (2003) proposed a possible role of sexual selection and sexual dimorphism under conditions of ecological destabilization. The multiple outcomes of this theoretical model seem to indicate that there is a conflict between speciation and dimorphism, i.e., sexual selection for increased dimorphism reduces the possibility of speciation. In this model, under conditions of changing ecological landscapes during the course of human evolution, some species undergo an adaptive speciation while others evolve alternatives to speciation, including an increase in sexual dimorphism. This may be a useful explanatory factor in the differing levels of sexual dimorphism observed in hominin taxa.
For the most part, although sexing is critical to almost all types of skeletal analysis, our ability to apply this to extinct members of the hominin lineage is difficult. Since the number of skeletons present in any analysis of the life history or paleodemography of fossil hominin samples is so small, it makes sense to pool both the male and female skeletons and consider them as a single sample.

Among the confounding difficulties which have led to the pooling of samples is the inability to distinguish dimorphic characteristics among the earliest hominins (including *Australopithecus*). Numerous arguments focused on the evolution of sex differences in the pelvis and in other features related to dimorphism (i.e., female vs. male stature). Assuming the obstetrical dilemma in human evolution, and without knowledge as to the exact time frame when this would have occurred in human biological history, research has pointed to possible evolutionary mechanisms responsible for these differences including sexual selection for more fecund female mates. Guegan et al. (2000) presented data on populations in 38 countries to show that stature dimorphism in females in modern populations could be correlated with complications associated with pregnancy and the birthing process. Nettle (2002) has shown that female height is correlated with reproductive success but that maximum success occurred in females who were below the mean for height. Integrating the data from both these studies suggests that over time female height may be more conservative than male height. Thus, differences in dimorphism in human evolution may be related in large measure to variations in males, with female biology remaining somewhat stable. Studies of living humans appear to show that females are less vulnerable to environmental influences on growth processes than males (Stini 1985).

Sexual selection has also been implicated in the evolution of body dimorphism. For example, Pawlowski and Grabarczyk (2003) have argued that the low center of gravity in females, manifest after puberty and adaptive for both pregnancy and the carrying of infants, is the result of sexual selection by males for this specific female body form.

The *os coxa* is the single most dimorphic feature of humans and, according to Tague and Lovejoy (1986), probably has been for at least the last few million years. The anatomy of the known sample of hominin pelvic bones appears to be related to the conflicting needs of both efficient bipedality and the problems associated with birthing relatively large-brained babies with broad shoulders. It appears that some of the dimorphism present in adult pelvises might actually be present *in utero* (Holcomb and Konigsberg 1995). In addition, it is possible that race affects the expression of sexual dimorphism in the pelvis (Patriquin et al. 2003). In modern populations, the accuracy of the assignation of sex using the pelvis is close to 95% (Murai et al. 1999; Bruzek 2002). According to Leutenegger (1982), all primate females possess a greater ischiopubic index than males do.
Although biases exist in the determination of sex based on features of the skull, there appear to be differential differences between the preservation of female versus male skeletons (Walker et al. 1988). There is a tendency to identify older female skulls as male. This is no doubt a direct consequence that the skull becomes more “masculine” as it ages (Meindl et al. 1985); it is difficult, however, to know if this bias applies to fossil hominin studies. Susanne et al. (1985) present data on age changes in cephalic dimensions.

The distinctive features of the male and female os coxae are universal in living humans (and probably most extinct species as well), although there might be some variation in the degree of expression of the differences within these populations. The same, however, does not apply to other characteristics used in sexing the skeleton. Thus, in concert with the techniques applied to determining the age at death of a skeleton, sexing techniques must be population specific. This can be accomplished either through analysis based on the data accumulated on a reference sample of known sex (Steyn and Iscan 1997; Graw et al. 1999; Mall et al. 2000; Asala 2001; Schiwy-Bochat 2001; Pettenati-Soubayroux et al. 2002; Bidmos and Asala 2004) or through seriation of a group of unknown specimens (White 2000) and the application of a variety of statistical techniques to the data. Safont et al. (2000) have employed this method on a modern human sample. Since it can generally be assumed that an unknown sample contains both male and female skeletons, it is reasonable to sort the skeletal elements into male and female categories.

In fossil hominin studies, in contrast, because it is not possible to derive a set of metric or nonmetric characteristics associated with sex (or age for that matter) on an applicable reference sample, it is necessary to derive a set a parameters that best describes (statistically or with other methods) the variation in the sample being examined. Depending on its place in the hominin timescale, comparisons can be made to hominoids, especially chimpanzees, or to living human populations.

An example of this type of analysis was that of Reno et al. (2003) on a fossil sample from Afar Locality A.L. 333 and assigned to A. afarensis (see also the commentary by Larson 2003). Since the A.L. 333 locale is considered on geological grounds to be a simultaneous death assemblage, it was assumed that the fossil sample represented males and females of the same species. Simulations were undertaken to compare the Afar fossils to modern humans, chimpanzees, and gorillas. Using measurements of postcranial elements, A. afarensis was found to be closest to H. sapiens in degree of sexual dimorphism (with H. sapiens intermediate between the monomorphism of chimps and extreme dimorphism in Gorilla in postcranial dimensions). These researchers also question the simple extrapolation of dimorphism to social categories in primates. They emphasized the striking
dissimilarity in sexual dimorphism between canine tooth maturation in chimpanzees in conjunction with their almost negligible postcranial dimorphism. They conclude that *A. afarensis* possessed a very distinct pattern from polygamous chimpanzees. In this analysis, monogamy is probably the most likely pattern of social organization among these early fossil hominins. Richmond and Jungers (1995) and Lockwood et al. (1996, 2000) have also investigated the pattern of sexual dimorphism in *A. afarensis*. Examinations of sexual dimorphism have been performed on *A. boisei* specimens (Silverman et al. 2001), face dimorphism in *A. africanus* (Lockwood 1999), sexual dimorphism in *Paranthropus robustus* and early *Homo* (Susman et al. 2001), and the mandibular metrics of a comparative sample of *H. sapiens, Pan troglodytes, Pongo pygmaeus pygmaeus*, and *Gorilla gorilla gorilla* (Humphrey et al. 1999).

One of the most extensive investigations of sex assignation in a specific sample was undertaken by Bermudez de Castro and Nicolas (1997) on the fossil materials from the Spanish site of Atapuerca (SH). A total sample of 32 individuals were examined, with a 1:1 sex ratio derived from the specimens that could be sexed. They report an overrepresentation of females in the age category between 16 and 20 years. This may be a reflection of a high female mortality in early child bearing years. Additional studies on this collection have been done by Rosas et al. (2002) and Bermudez de Castro et al. (2001), who examined the level of sexual dimorphism in the sample as revealed in mandibular and dental measurements, respectively. Bermudez de Castro et al. (2001) concluded that the dentition showed greater sexual dimorphism than that found in modern humans. In the mandible, sexual dimorphic patterns were present but differed from those in modern humans. Arsuaga et al. (1997) looked at body size and cranial capacity dimorphism, reporting a degree of dimorphism similar to other Middle Pleistocene hominins and modern humans.

Wolpoff (1999) noted that sexual dimorphic characteristics, including overall cranial size and capacity, vault thickness, superstructures and toruses (mastoid process, sagittal keel, nuchal torus), forehead curvature, and functionally related features of facial size and robustness, vary in expression in recent human geographic populations and are a reflection of the hominin evolutionary past in each region. On the basis of these observations, Wolpoff (1999) identified male and female sex differences in an extensive sample of Middle and Upper Pleistocene hominins. Weidenreich’s (1935, 1943) detailed studies of the morphology of the *H. erectus* sample from Zhoukoudian suggested a relatively low level of sexual dimorphism, a conclusion that was also found by Mann (1981), who studied the virtually complete Zhoukoudian cast collection.

Other studies of the application of techniques of sexing to fossil hominin materials include the work of Coqueugniot et al. (2000) as well as the application
of mandibular ramus posterior flexure to a sample of Neandertals and early modern humans (Loth and Henneberg 1996); dimorphism in chin morphology in the specimens from Klasies River Mouth (Lam et al. 1996); an analysis of the Kebara 2 pelvis using various bony dimensions (Tague 1992). Finally, Trinkaus (1980) employed indices of robusticity of postcranial elements to show that European and Near Eastern Neandertal limb bones show the same degree of sexual dimorphism as modern humans in this feature.

22.3 Other paleodemographic parameters

22.3.1 Life history

Life history variables, including brain and body size, neonatal weight, length of gestation, life span, age of sexual maturation, age of weaning, are all correlated to each other; various theories have been proposed to explain the evolutionary mechanisms through which this occurs. Using data on living primates, Harvey and Clutton-Brock (1985) noted that the distinctive features of this mammalian order are large brains, prolonged maturation, and long life spans. Understanding the relationships of these variables to each other from an evolutionary perspective has been the subject of numerous projects. Pereira (1993) has summarized this work, including a discussion of allometric or correlation analysis. Heterochrony has also been used to explain life history developmental processes that may have characterized human evolution (Gould 1977; Minugh-Purvis and McNamara 2002). The usefulness of heterochrony has been critically evaluated by Godfrey and Sutherland (1996) and Shea (1989).

Gage (1998) compiled all known mortality and fertility data on a broad group of primates. He concluded that a comparison of human and chimpanzees life histories indicates that delayed maturation is a characteristic of humans, whereas an extension of the overall time frame of reproduction, along with an increase in the calculated rate of aging, appears characteristic of chimpanzees. Chimpanzee character state of life history is evolutionarily derived and thus does not make a useful model for understanding the last common ancestor of chimpanzees and humans. Although this observation must be kept in mind, chimpanzees are our closest living relative; the use of the growing body of literature on chimpanzee studies, especially in the wild, can reasonably provide a point of comparison not only to living humans but also to extinct members of the human lineage. The human life history pattern as described by Smith and Tompkins (1995 p 260) is that “humans take about twice as long to erupt teeth, twice as long to reach adulthood, and live about twice as long as great apes.” This probably
represents an overstatement of the unique position of humans within the Primate Order. For example, the work of Zihlman et al. (2004) shows that wild chimpanzees take upward of 3 years longer to mature than animals in captivity. Using M1 eruption times as a proxy for other maturational events, chimpanzees now appear to overlap the range of modern humans. Recent work on sexual maturation of chimpanzees in the wild shows that some groups do not mature until 14 years of age (Sugiyama 2004b).

A comprehensive description of life history in mammals was analyzed using an energetics model by Hill and Kaplan (1999), and an evolutionary model of longevity increases was presented by Carey and Judge (2001). Using game theory, Brommer (2000) also discussed the evolution of life history using the evolutionary concept of fitness. Using seven anthropoid primate species, Leigh (2004) has shown that life history among the primates cannot be related solely to brain size but more specifically to the point in life history in which brain growth occurs. The correlation of diet to primate life history has been addressed by Godfrey et al. (2001) and Kaplan et al. (2000). Day et al. (2002) associated life history variables across various species of animals with the level of predator pressure. Bolnick and Doebeli (2003) have also proposed that species confronted with ecological instability adapt in a number of ways, including changes in sexual dimorphism, speciation, or an alteration of ontogenetic processes that reduces the influence of the destabilizing ecological events. It is possible that some and all combinations of these outcomes have occurred in human evolution. Of critical importance is the plausible role of ontogenetic changes as an adaptive feature in at least some species in human evolution. If, for example, A. afarensis does indeed show little sexual dimorphism and, by extrapolation from primate models, this indicates a pattern of monogamy among these hominins, then it is indeed possible that altricial young were also part of this pattern, further supporting Lovejoy’s (1981) hypothesis.

Since life history theory is built on the foundation of living animals, it becomes a challenge to integrate fossil hominin species into these analyses. It has been argued that correlations of brain and body size to life history variables supersede any possible analysis of the fossil materials and that all data must be interpreted in this context. Certainly evolutionary events have produced unique combinations of life history features in the past; and if it is possible to integrate these into the correlation studies, it might serve to more fully represent the life history of hominins across time and space. There appears to be no consensus as to when in the course of hominin evolution modern human-like life histories evolved. Various perspectives on the evolution of life history have been presented (Smith and Thompkins 1995; Tillier 1995; Bogin 1999; Minugh-Purvis and McNamara 2002; Thompson et al. 2003).
In order to reconstruct the life history patterns of earlier hominins, most research has concentrated on the examination of those biological patterns that may yield data documenting rates of maturation. Overwhelmingly, this has meant focusing on the developing dentition with the assumption that the maturation of dental hard tissues and the sequential eruption of the teeth are under strong genetic control, with a minimum of possible environmental influence. Unlike bone, which is subject to remodeling throughout life, dental enamel appears to be relatively stable, preserving the initial structures. Most research on the maturation patterns of earlier hominins focuses either on the pattern and sequential eruption of the developing dentition or on the examination of dental microstructures as time-dependent features. Comparisons between fossil hominins are most usually made to the living great apes, especially chimpanzees, and to modern humans.

### 22.3.1.1 Patterns of maturation

Chimpanzees and humans are characterized by different life history strategies; this should be apparent in the dental calcification and eruption schedules of the two species. What if any pattern differences have been described? The most prominent is that chimpanzees have an accelerated development and eruption of the M1 especially in relationship to the development of I1 (Smith 1994, 2000; critique by Mann et al. 1990). It has also been observed that chimpanzees have a time overlap in the calcification schedule of the molar sequence (Anemone 2002). And finally, chimpanzees have a significant difference in the pattern of calcification and eruption of the permanent canines. Since this tooth is so functionally and anatomically distinct from that of humans, the canine is probably not a good indicator of life history pattern differences (Lampl et al. 1993).

Various models have been used to explain these differences. These include differences in life history (see for example Smith 2000 and the critique by Lampl et al. 2000) but other explanations have been proposed (Simpson et al. 1990, Simpson and Kunos 1998).

Many scholars have undertaken general discussions of how dental development of fossil hominin samples can be understood as part of life history (Bogin 1990, 1999; Smith and Tompkins 1995; Bogin and Smith 1996; Kuykendall 2003). Discussions of life history in the early phases of human evolution have been extensively discussed (Dean 1985; Smith 1986, 1989, 1991, 1992, 1994; Conroy and Vannier 1987, 1989; Mann et al. 1987; Conroy 1988; Mann 1988; Wolpoff et al. 1988; Conroy and Kuykendall 1995; Smith and Tompkins 1995; Clegg and Aiello 1999; Dean et al. 2001; Moggi-Cecchi 2001). Many investigations have also focused on the later phases of human evolution (Dean et al. 1986; Tompkins

22.3.1.2 Dental microstructure

Some researchers have maintained that dental enamel and dentin microstructures are useful tools for the reconstruction of the timing of dental development. Dental microstructure is a hard tissue relic of the ontogenetic processes that govern tooth growth (Ramirez-Rozzi 1998a, b, 2002; Ramirez-Rozzi et al. 1999; Ramirez Rozzi and Bermudez de Castro 2004; Guatelli-Steinberg et al. 2005). Conversely, it is also possible that these structures reflect a microworld of minianatomical detail that may or may not be useful in the timing of dental development or the reconstruction of phylogenetic relationships.

Bromage and Dean (1985) presented data to support the idea that the timing of tooth development in fossil hominins could be estimated using microstructural details of enamel. Since then there has been much debate about the applicability of these hard tissues to understanding of the times of maturational events in earlier hominins (Bromage and Dean 1985; Dean 1985, 1987a–c, 1998, 1999, 2000a, b; Beynon and Dean 1987, 1988; Beynon and Wood 1987; Bromage 1987, 1989, 1990; Mann et al. 1987, 1990; Dean et al. 1993; Beynon et al. 1998; Moggi-Cecchi et al. 1998; Dean and Reid 2001; Moggi-Cecchi 2001; Ramirez Rozzi and Bermudez de Castro 2004; Guatelli-Steinberg et al. 2005).

Finally, growth and development studies have been undertaken on the femur (Tardieu 1998) and pelvis (Berge 1995) of fossil hominins. In both these studies, although a humanlike pattern of growth best describes all members of the genus Homo, australopithecine-grade hominins probably had growth trajectories similar to those in chimpanzees.

What can be concluded from this diverse body of research? Even where a virtually complete skeleton of a youngster (KNM-WT 15000) has been recovered, it has been difficult to reach a consensus about its rate of maturation and its age at death (Clegg and Aiello 1999). The diversity of research on the patterns of maturation in earlier hominins is well illustrated in the edited volume, *Patterns of Growth and Development in the Genus Homo* (Thompson et al. 2003). But it remains very difficult to compare the work presented by one researcher with that of another because the fossil samples and the techniques of analysis that are employed are so different.

However, with some reservations, the research results appear to support the general recognition that, by the appearance of members of the genus Homo, there was a maturational shift to a more modern humanlike pattern, although that pattern might be manifested in slightly different ways in individual species within...
Homo. If this cautious conclusion continues to be supported by additional data, it will justify the use of the growth and development patterns of modern humans as the reference data in paleodemographic reconstructions. What sorts of patterns characterized the maturation of the still earlier members of the hominin lineage remains uncertain. However, recent information documenting significant variation in maturational events in both modern humans and chimpanzees suggests that the traditional notion of a uniquely human prolonged period of maturation may have to be reconsidered.

22.4 Conclusions

1. The paleodemography of extinct hominin taxa has not produced a corpus of dependable data, and it remains possible that reasonable population-based demographic parameters from fossil assemblages will remain unattainable into the foreseeable future.

2. In contrast, life history reconstruction, establishing some basic parameters associated with demographic analysis, is a more achievable goal. Among these parameters are potential life expectancy, time of maturation, and age at weaning. The accumulation of additional information on living primates has made the simple association of life history variables with gross size variables (e.g., brain size and body size) increasingly untenable.

3. The larger the database on captive and wild chimpanzees, the more difficult it becomes to clearly distinguish their life history variables from those of living humans. This will certainly affect our interpretations of life history in fossil hominin species. Similarly, the more knowledge we have of the extent of modern human variation and plasticity, the more difficult it becomes to specifically categorize human life history.

4. Culture (in conjunction with significant developmental plasticity) may be more important as regard understanding human life history than biology, especially in variations in the expression of features such as longevity and age at weaning, and perhaps even maturation rates, general reduction in the age of first menstruation, and others. Such factors may better explain the large ranges of variation that have been observed in both humans and chimpanzees.

In sum, we know very little of the demography of earlier members of the hominin lineage, and much that appears to be understood requires reevaluation on the basis of newly acquired data from nonhuman primates and the increasing comprehension of the range of variation that exists in human growth and maturation.


Gunter-Stevenberg D, Reid DJ, Bishop TA, Larson CS (2005) Anterior tooth growth periods in Neandertals were comparable to those of modern humans. Proc Natl Acad Sci 102: 14197–14202
Mann AE (1968) The paleodemography of Australopithecines. Ph.D. Thesis in Anthropology, University of California, Berkeley
Mann AE (1975) Paleodemographic aspects of the South African Australopithecines. University of Pennsylvania, Department of Anthropology, Philadelphia
Paine RR (2000) If a population crashes in prehistory, and there is no paleodemographer there to hear it, does it make a sound? Am J Phys Anthropol 112: 181–190
Ramirez-Rozzi F (1998a) Can enamel microstructure be used to establish the presence of different species of Plio-Pleistocene hominids from Omo, Ethiopia? J Hum Evol 35: 543–576
Reno PL, Meindl RS, McCollum MA, Lovejoy CO (2003) Sexual dimorphism in Australopithecus australiensis was similar to that of modern humans. Proc Natl Acad Sci 100: 9404–9409


Weiss KM (1973) Demographic models for anthropology. Memoirs of the Society for American Archaeology, Number 27, American Antiquity 38, Number 2, Part 2
23 Modeling the Past: The Primatological Approach

R. W. Sussman · Donna Hart

Abstract

Many models have been developed to depict the behavior and ecology of our earliest relatives. However, the Man the Hunter model has been the most widely accepted way of viewing human evolution. This theory gained ground in the mid-twentieth century and has been recycled ever since under various guises in the scientific and popular literature. Many human traits, such as bipedalism, monogamy, territoriality, tool use, technological invention, male aggression, group-living, and sociality, are often linked to this perspective. However, while theories and associations of human aggressive hunters abound, they are rarely based on the two evidentiary approaches that shed light on early hominid ecology and behavior—living primate models and the fossil record. Here, we outline a methodology of reconstructing early human behavior by using both the fossil record and extant primate ecology and behavior. Data on early human fossils, on modern primates living in similar habitats to our earliest ancestors, and on predators both today and in the distant past indicate that Man the Hunted may be a more accurate descriptor of our earliest relatives. We describe the evidence for the Man the Hunted theory, some of the behavioral patterns that were needed to protect our earliest ancestors from predation, and how this may lead to a new perspective on certain aspects of human nature.

23.1 Introduction

In the early 1950s, Bartholomew and Birdsell (1953) published their classic paper *Ecology and the Protohominids*. This was among the first attempts to develop a method for reconstructing the behavior and ecology of our earliest hominid ancestors. They suggested that it should “be possible to extrapolate upward from ecological data on other mammals and suggest the biological attributes of the protohominids and to extrapolate downwards from ethnological data on hunting and collecting peoples and suggest the minimal cultural attributes of the protohominids” (p 481). This paper was precipitated by the realization, in the
late 1940s and early 1950s, that australopithecines were indeed our earliest ancestors and that in most ways they were more similar to nonhuman primates than to modern humans (Sussman 2000). Since then a number of models have been proposed using research on nonhuman primates to reconstruct early human ecology and behavior.

The most persistent model and one that occurs repeatedly in the literature on human evolution is the “Man the Hunter” model. Among the earliest renditions of this scenario was that developed by Dart after his fossil discoveries in South Africa. From the fragmented state of the bones and the damage on the early hominid skulls, he concluded that this species had used bone, tooth, and antler tools to kill, butcher, and eat prey, as well as to kill one another (Dart 1953). Between 1961 and 1976, Dart’s view was popularized by Robert Ardrey in a number of books, including African Genesis, The Hunting Hypothesis, and The Territorial Imperative. The evidence for Dart’s theory was not good, however. On examination of the same evidence, Brain (1981) noted that the bones associated with these early hominids were exactly like fragments left by leopards and hyenas. It seems that Dart’s australopithecines may have been the hunted and not the hunters.

In the late 1960s, Washburn and his colleagues restated the Man the Hunter theme (Washburn and Lancaster 1968). Rather than amassing an adequate amount of evidence to prove their theory, Washburn and his colleagues used the nineteenth-century concept of cultural “survivals.” Behaviors like hunting that persist from an earlier time but are no longer useful in modern society were seen as part of our inherent biology. Another major proponent of the Man the Hunter model was E.O. Wilson in the mid-1970s. He claimed that a number of human traits, such as territoriality, aggressive dominance hierarchies, male dominance over females, permanent male–female bonds, and female philopatry, are relatively constant among our primate relatives and persist throughout human evolution and in human societies, generally. Furthermore, he believed these traits provide important evidence of our hunting past (Wilson 1975).

One of the most recent models of our hunting past is that provided by Wrangham and Peterson (1996). They proposed that killing and violence against adults of their own species are traits shared by hominids and chimpanzees and that, since these traits are rare among other animals, they were inherited from our ancient common ancestor. In fact, unlike earlier models which proposed that killing developed out of our propensity for hunting, in Wrangham’s and Peterson’s model it is this violent nature and natural blood lust that makes both humans and chimpanzees such good hunters. The latter three models have been reviewed elsewhere and found to be lacking in substantive evidence (Sussman 1999a, b, 2004; Marks 2002; Hart and Sussman 2005; Fry 2006).
23.2 Reconstructing the behavior of the earliest hominids

If the evidence for Man the Hunter has not been good, what kinds of evidence should be used in attempts to reconstruct the behavior of our earliest ancestors?

The most important evidence includes careful examination and understanding of the actual skeletal remains of the creatures. Other fossil evidence, such as tools or footprints, left by our earliest relatives and fossil materials that give us clues about the environment in which they lived, such as fauna, flora, or water sources, also must be examined. These fossils provide the most important data for accurate reconstructions. Many Man the Hunter theories fall short in the critical examination of the fossil evidence (if fossils are used at all). Stanford noted this lack of fossil evidence in such evolutionary theories: “… models of hominid evolution are, appearances notwithstanding, virtually fossil-free” (Allen and Stanford 1992, Stanford personal communication).

Besides fossils, other types of secondary evidence used in reconstructions are less reliable but, nonetheless, can offer insights. We would rank these in the following order as far as applicability to reconstructing early hominid lifestyles: (1) the behavior of nonhuman primates living under similar ecological conditions to those of our early ancestors. It is best to keep timing in mind with this approach since habitats and climates change throughout our evolutionary history. (2) The behavior of our genetically closest primate relatives such as chimpanzees, bonobos, and gorillas. However, lumping all the great apes together as one analog when they are so diverse is dangerous. Although some characteristics may remain conservative within a taxonomic group, others are not shared. For example, monogamous pair bonds among the lesser apes or upright posture among the apes might be considered phylogenetically conservative traits, shared by all or most species within a taxon. (3) Characteristics shared by certain (or all) modern humans that might also be similar to our earliest ancestors. Modern foragers, however, are just as advanced and evolved within their own cultures and environments as any Western urban dwellers. Our least confident recommendation is (4) the behavior of other animal species that might be living under similar conditions or share some aspects of the lifestyle of early humans such as certain carnivore or prey species.

In using any of these types of secondary evidence, if we are not extremely careful (because in many cases similar looking behaviors are not the same), we can end up comparing apples with oranges or lions with hominids. Obviously, words with loaded meaning for humans—war, rape, murder, infanticide, and genocide to name a few—must be used with extreme caution when referring to
the activities of nonhuman species. In this regard, Marks (2002 p 104) warns against “... a science of metaphorical, not of biological, connections.”

We cannot, therefore, necessarily impute correlation between human ancestors and data based on extant carnivores, modern human foragers, or great apes. For example, even the concept of hunting in chimps and humans is quite different. Present-day human hunters purposely search for animal prey but chimpanzees do not: “Instead, they forage for plant foods and eat prey animals opportunistically in the course of looking for fruits and leaves” (Stanford 1999 p 48). Furthermore, reconstructions must always be compatible with the actual fossil data—the fossils are real but the models we construct are hypothetical and must constantly be tested and reconfirmed. Lastly, when attempting to construct models of our early ancestors’ behavior, it is necessary to be precise about timing. If we say our earliest human ancestors (those who lived 7 Ma) behaved in a certain way, we cannot use fossil evidence from 2 Ma.

As a case in point about timing, we pose the question: Could hunting have occurred without tools? The first evidence of stone tools comes from around 2.6 Ma (Semaw 2000). The earliest hominin fossils, however, date from almost 7 Ma, at least 4 Myr before the first stone tools. When we look at the fossil evidence, hunting came quite late to our human family. Interpretations of hominin behavior, therefore, should be conservative and cautious, as stated by Klein (1999 p 306): “... the mere presence of animal bones at archeological sites does not prove that hominids were killing animals or even necessarily exploiting meat. Indeed, as was the case in the earlier South African sites, the hominin remains themselves may have been the meal refuse of large carnivores.”

The transition to hunting as a dominant way of life does not appear to have started until after the appearance of our own genus, Homo, and may not have even begun with the earliest members of our genus. Homo erectus has been given credit in the past for existing as a large animal hunter, and dates as far back as 1.75 Ma have been hypothesized for such a lifestyle. But if we take a conservative approach to this subject—looking only at facts and fossils and not imaginative speculations—the first unambiguous indications of hunting may be amazingly recent (Binford 1992; Klein 1999).

No hard archeological evidence, in other words, no fossil evidence of tools designed for hunting exist earlier in time than a finely shaped wooden spear excavated at Schöningen, Germany, dated at approximately 400,000 years of age (Dennell 1997; Theime 1997). The famous Torralba and Ambrona sites in Spain, dated at 500 ka, contain huge numbers of large mammal bones and were thought to represent unquestionable evidence of megafauna killed by Pleistocene hunters. Now these two sites are being reconsidered in the light of better archeological analysis. Elephant bones at these sites could just as likely
represent natural deaths or carnivore kills as the remains of human hunting (Klein 1987, 1999).

Further, no hominids were large-scale hunters before they had the use of fire (because of their dentition and alimentary tract, points we will elucidate later), although insects, small vertebrates, lizards, and birds were likely eaten opportunistically. The best evidence for the controlled use of fire appears around 800 ka in Israel (Goren-Inbar et al. 2004). Klein (1999 p 160) states: “The assumption of consistent hunting has been challenged, especially by archeologists who argue that the evidence does not prove the hunting hypothesis . . . it is crucial to remember (although not as exciting) that probably the majority of calories [came] from gathering plant foods.”

### 23.3 Dentition and diet

Whether *H. erectus* or any other hominid hunted or scavenged before 800 ka may be a moot question. Hunting would only be an activity undertaken if early hunters could eat what they killed, and to eat meat it is necessary to have teeth capable of masticating and processing meat.

Man the Hunter models of human evolution assume that a significant portion of our earliest ancestors’ diets must have come from killing and eating meat from relatively large mammals. By comparing the characteristics of the dental and jaw morphology of various living primates with those of fossils, we can make inferences about the diets of early hominids. Teaford and Ungar (2000; Ungar 2004) carried out just such a comparison in an attempt to reconstruct the early hominid diet. Using such features as tooth size, tooth shape, enamel structure, dental microwear, jaw biomechanics, and occlusal slope and relief of the lower molars, they found that the earliest humans had a unique combination of dental characteristics and a diet different from modern apes or modern humans.

*Australopithecus afarensis* is characterized by jawbones that are thick, with relatively small incisors and canines in relation to molars. The molars, by comparison with other primates, are huge, flat and blunt, show less slope and relief, and lack the long-shearing crests necessary to mince flesh. *A. afarensis* also had larger front than back molars. The dental enamel is thick and microwear on the teeth is a mosaic of gorilla-like fine-wear striations (indicating leaf-eating) and baboon-like pits and microflakes (indicating fruits, seeds, and tubers in the diet).

When mammalian omnivore species are compared, body mass is positively correlated with inclusion of plant protein in lieu of animal protein in the diet. If insects are ingested, they are usually social ones that come in large “packages”
because it takes a higher amount of energy to capture individual insects. In studies of mid- to large-sized primates, such as macaques, baboons, chimpanzees, and modern human foragers, in which the amount of time spent obtaining animal protein has been quantified, the total is very low, usually making up less than 5% of the diet (Garber 1987; Sussman 1999c).

Given these facts, we hypothesize that early humans were able to exploit a wide range of dietary resources, including hard, brittle foods (tough fruits, nuts, seeds, and pods), as well as soft, weak foods (ripe fruits, young leaves and herbs, flowers and buds). They may also have been able to eat abrasive objects, including gritty plant parts such as grass seeds, roots, rhizomes, and underground tubers. As stated by Teaford and Ungar (2000 pp 13508–13509), “this ability to eat both hard and soft foods, plus abrasive and nonabrasive foods, would have left early hominids particularly well suited for life in a variety of habitats, ranging from gallery forest to open savanna.” Dental morphology indicates that the earliest hominids would have had difficulty breaking down tough pliant plant foods, such as soft, fibrous seed coats, and the veins and stems of mature leaves. Another tough plant food that our early ancestors would have had difficulty processing was meat. Teaford and Ungar (2000 p 13509) stress that “the early hominids were not dentally preadapted to eat meat — they simply did not have the sharp, reciprocally concave shearing blades necessary to retain and cut such foods.”

Both modern chimpanzees and humans have an alimentary track that is neither specialized for eating leaves nor animal protein: instead, it is more generalized, similar to the majority of primates who are omnivorous and eat a mixture of food types (Chivers and Hladik 1980; Martin et al. 1985; Martin 1990). Furthermore, at least two-thirds of modern-day foragers’ food comes from women gathering plant foods and, in the process, opportunistically capturing small mammals and reptiles. Less than one-third of the diet (the meat portion brought in by dedicated male hunters) serves to supplement their foraged nutritional intake, except in cold climates or where fishing is prevalent (Marlowe 2005).

Modern dietary concerns in industrial societies revolve around the amounts of both fat and red meat consumed by the average person. The medical community warns that meat should be ingested in limited quantities. “Diseases of affluence” caused by high-protein, high-fat diets include raised cholesterol levels, high blood pressure, heart disease, stroke, breast cancer, colon cancer, and diabetes—all correlated to a diet exorbitantly rich in red meat. With colon cancer, in particular, startling data are available: Daily red-meat eaters are two and one-half times more likely to develop this cancer than people adhering to a mostly vegetarian diet (Willett et al. 1990). In a study in which Chinese rural villagers (eating a traditional diet low in meat) were compared to their urban counterparts
(who ate more meat), it was found that individuals following a low-fat, low-meat diet suffered less anemia and osteoporosis (conditions commonly associated with food low in animal products) than individuals higher on the meat-consuming ladder (Chen et al. 1990; Campbell and Campbell 2005). Campbell states: “We’re basically a vegetarian species and should be eating a wide variety of plant foods and minimizing our intake of animal foods” (Brody 1990 p C2).

No hominids could have hunted for a meat diet on a large scale before the advent of controlled fire. We did not have the dentition or the digestive tract of a carnivore. The anatomy and physiology of early hominids did not particularly suit them for digesting meat until the mastery of cooking. The human intestinal tract is short and predigestion of meat by fire had to precede any major meat eating, although we still require certain nutrients that are not obtainable from a meat diet. As stated earlier, the oldest known hearths with good evidence for controlled use of fire are dated at approximately 800,000 years.

### 23.4 Locomotion

By far the best known of early australopithecine species is *A. afarensis*, with many fossil remains dating from 3.6 to 2.9 Ma and possibly as far back as 5 Ma. Collections from Hadar, Ethiopia alone have yielded 250 specimens, representing at least 35 individuals, and there are a number of other East African digs that contain remains of this species. Specimens include Lucy (dated at 3.2 Ma), the most complete adult skeleton from this time period, and fossil footprints from Laetoli, Tanzania ash deposits (dated at 3.6 Ma). Furthermore, most hypotheses concerning human evolution position *A. afarensis* as a pivotal species from which all other later hominids, including *Homo*, evolved (Fleagle 1999; Conroy 2005).

Given the previous facts, we see *A. afarensis* as a good species to examine when attempting to reconstruct the appearance and behavior of one of our early human ancestors.

Terrestrial bipedalism is a hallmark of the whole fossil hominid family. This mode of locomotion can be inferred from fossil specimens nearly 7-Myr old (Galik et al. 2004). It appears long before the vast growth of open grasslands in Africa and before the expansion of human brain size and recognizable stone tool making. Besides fossilized bones, direct evidence of early bipedalism comes from the Laetoli footprints where two hominids were walking together in soft ash almost 4 Ma (White 1980; Day 1985; Tuttle 1985; Klein 1999). However, looking at the skeletal evidence, the locomotion of these early hominids was not exactly identical to ours. In fact, *A. afarensis* seems to have been a primate equally at home in the trees or on the ground.
This is indicated by a number of factors. First, the limb proportions are different from modern humans. The arms are similar in proportion to modern humans, but the legs are relatively shorter—more apelike—which implies the use of suspensory locomotion in the trees (Kimbel et al. 1994). Other aspects of the upper limbs also retain a number of features indicating an ability to move easily in the trees. The wrist and hand bones are quite chimpanzee-like; the finger and toe bones are slender and curved as in apes, giving *A. afarensis* grasping capabilities compatible with suspensory behaviors. The toe bones are relatively longer and more curved than in *H. sapiens*, the joints of the hands and feet, and the overall proportions of the foot bones all reinforce evidence for climbing adaptations and arboreal activity. Nevertheless, the relative thumb length of these hominids is closer to that of modern humans than it is to chimpanzees (Susman et al. 1984; Smith 1995; Corkern 1997; Alba et al. 2003; Conroy 2005).

The pelvis and lower limbs of *A. afarensis* are a mixture of humanlike and apelike features. These anatomical components and the shorter leg length indicate that *A. afarensis* may have used less energy while walking, whereas transition speeds from walking to running may have been lower with slower running speeds than modern longer legged humans (Steudel-Numbers 2003). Overall, Rak (1991 p 283) summarizes: “Although clearly bipedal and highly terrestrial, Lucy evidently achieved this mode of locomotion through a solution of her own.”

Does this mixture of anatomical features indicate anything about the efficiency of early hominid locomotion? It was once thought that the Neotropical marmosets and tamarins were restricted in their ability to move on small branches because they have claws instead of the standard primate nails on their hands and feet. However, claws do not restrict callitrichid locomotion on thin branches; their claws also enable them to utilize large trunks much like squirrels do. Claws allow them to be more versatile, and they can use a wider range of arboreal habitats than most other New World monkeys (Garber 1984; Sussman and Kinzey 1984). It appears that the combination of skeletal characteristics found in *A. afarensis* enabled this hominid to be versatile in a similar way. They were able to use the ground and the trees equally and successfully for a very long time. We believe these early hominids were well-adapted to their environment and not in the least inhibited by switching back and forth from bipedalism on the ground to quadrupedalism in the trees.

There are at least seven different “models” that have been proposed to account for bipedalism as a hominid adaptive strategy: the *tool using and making* model (Darwin 1874; Washburn 1960; Sinclair et al. 1986), the *carrying* model (Hewes 1961, 1964; Lovejoy 1981), the *vigilance* model (Dart and Craig 1959; Day 1986), the *heat dissipation* model (Wheeler 1984, 1991), the *energy efficiency*
model (Rodman and McHenry 1980; Steudel 1994), the display model (Jablonski and Chaplin 1993), and the foraging model (Hunt 1994). However, there are many primates who spend most of their time on the ground and none of these developed bipedalism, even though each of the theorized advantages presumably also would have accrued to them.

It is difficult to separate consequence from causation. None of the theories may be causative; instead, all the theoretical causes may be results of a primate preadaptation to being bipedal. All the great apes are preadapted to bipedality. When our ancestors came down from the trees, bipedalism was possible because of body proportions and suspensory adaptations—longer arms and shorter legs that allow gibbons, orangutans, and chimpanzees to hang from trees and forage for fruit. All apes have varying capacities for erect posture and are able to walk upright for short periods of time; bonobos, especially, will stride upright with humanlike posture. However, we propose that when the earliest hominids began using the ground for a major portion of their activities, their body proportions were more suited for bipedalism than other forms of locomotion, i.e., quadrupedalism or knuckle-walking.

Although it is important to see early hominids in the context of hominid evolution, it is equally important to realize that in the same way that they were not little people, they also were not just bipedal chimps, but the beginning of a new radiation of very different hominoids. It is this uniqueness that makes reconstructing hominid origins so difficult. Thus although early hominids and their bipedal adaptations are certainly derived from an African apelike ancestry, human bipedalism is morphologically and physiologically different from the occasional facultative bipedal behaviors occasionally seen in other primates. The morphological and behavioral commitment to bipedalism that characterized early hominids suggests unique ecological and historical circumstances as well (Fleagle 1999 p 528).

Some species of primates are intrinsically adapted to edge habitats and are, therefore, able to take advantage of changing environments. We hypothesize that the earliest hominids were edge species (see later) and that they exploited a terrestrial habitat due to a developing mosaic environment that included climate change. Rather than seeking the factors that caused early human ancestors to become bipedal, we propose that it was a preadaptation that already existed, and it was efficient in a new habitat; the successes or added advantages were simply a byproduct. Tattersall (2003) has arrived at a similar conclusion.

Besides bipedalism and limb use, there are also solid conjectures of what our earliest ancestors were like as far as body build, height, weight, and brain capacity. From an examination of the skeleton of Lucy and various A. afarensis specimens,
it seems there was a considerable size difference between males and females. Although the canines of both sexes were relatively small and not dagger-like, they were larger and longer in males than females. The range of body size for *A. afarensis* individuals is estimated to be 30–45 kg (Fleagle 1999); the height of adults has been estimated at 1.0–1.7 m (Klein 1999). Lucy, for example, stood slightly over 1-m tall and weighed around 30 kg (Conroy 2005). If these estimates are accurate, we can extrapolate that female *A. afarensis* were the size of male baboons and males were the size of female chimpanzees.

The cranial capacity of these hominids is estimated at 400–500 cc—about the size of a modern chimpanzee but twice as large as Miocene fossil apes. On average, australopithecines and modern chimpanzees have brains that are two to three times larger than similar-sized mammals, whereas in modern humans brain size is six to seven times larger than other mammals. Using the encephalization quotient (brain size relative to body size) as a measurement, the brain of *A. afarensis* was slightly larger in relation to its body than that of modern chimpanzees (EQ = 2.4 for *A. afarensis* versus 2.0 for chimpanzees) (Boaz 1997). Thus, our ancestors were mid- to relatively large-sized primates with brains slightly larger than any living nonhuman primate, although only a fraction bigger than modern chimpanzees.

There are many speculations about the external appearance of australopithecines. Johanson and Edey (1981) reconstructed Lucy as follows: small but extremely powerful; bones robust for her size; heavily-muscled; arms longer in proportion to her legs and trunk than modern humans; hands like modern humans but fingers curled more when in a relaxed position; head more apelike than humanlike with prognathic jaws and no chin; possibly more body hair in comparison to modern humans.

### 23.5 Habitat of our earliest ancestor

Although many early theories on the evolution of our earliest ancestors stress the importance of arid, savanna environments, these do not seem to be the primary habitats, according to the fossil record, until after 2 Ma. The African climate became more arid in the time between 12 and 5 Ma, and equatorial forests were undoubtedly shrinking (Conroy 2005). However, the process that led to this climatic phenomenon also greatly enlarged areas of transitional zones between forest and adjacent savanna. Closed woodland forests were still widespread in East Africa 3.5 Ma, whereas the proportion of dry shrub to grassland habitats began to increase around 1.8 Ma (de Menocal and Bloemendal 1995; Schekleton 1995).
It is in these transitional zones that behavioral and anatomical changes may have been initiated in early hominid evolution. The flora and fauna remains that are found in association with fossil hominids of this time period indicate a mixed, mosaic environment, ecologically diverse and subject to seasonal and yearly changes in vegetation (Potts 1996; Wolpoff 1998; Conroy 2005). These environments were wetter than those in which later fossil hominids are found, and most fossil sites of this early time period contained some type of water source such as rivers and lakes (White et al. 1994; Wolde Gabriel et al. 1994). For example, at Hadar in Ethiopia the mammalian fauna remains suggest that a lake existed, surrounded by marshy environments fed by rivers flowing off the Ethiopian escarpment (Conroy 2005). A mosaic of habitats existed at Hadar consisting of closed and open woodland, bushland, and grassland.

The earliest hominids appear to be associated with variegated fringe environments or edges between forest and grassland. These types of habitats usually contain animal and plant species of both the forest and the grassland, as well as species unique to the borders between the two. The species adapted to these transitional habitats are often referred to as edge species. During these earliest times, it appears that hominids began to take advantage of the growing fringe environments, lessening competition with their sibling ape species which were better adapted to exploit the dense forest, and thus partitioning the niche occupied by the parent species of both apes and hominids into two narrower and less overlapping adaptive zones (Klein 1999; Conroy 2005).

From available evidence, we speculate that our early ancestors were able to exploit a great variety of food resources but were mainly fruit-eaters, probably supplementing this diet with some young leaves and other plant parts, social insects, and a small amount of opportunistically captured small vertebrate prey—lizards, small snakes, birds, and mammals.

Several other species of primates are intrinsically adapted to edge habitats and are also able to take advantage of changing environments. Ring-tailed lemurs in Madagascar, African baboons and vervet monkeys, and some Asian macaques and langurs are nonhuman primate examples. These, not coincidentally, are some of the most common and numerous of all living primates other than humans. The macaque genus, for example, has the widest geographical distribution of any nonhuman primate in Asia. Many macaque species in Asia are endangered, but the ones that have the healthiest populations (e.g., long-tailed macaques, *Macaca fascicularis*, and rhesus macaques, *Macaca mulatta*) are edge-adapted.

Certain habitats may tend to impose specific behavioral repertoires. Many argue that the more similar the DNA, the more similar will be the behaviors between related species. In that case, chimps and bonobos might be the best prototype for early human ancestors. However, if ecology is paramount, then
chimps and bonobos may be less suitable prototypes (although some traits between these close relatives may still be important and phylogenetically conservative) and edge species may be the best models for early humans. Fox (1967 p 419) stated: “... the problem of taking the great apes as models lies in the fact of their forest ecologies. Most modern students of primate evolution agree that we should pay close attention to ecology in order to understand the selection pressures at work on the evolving primate lines. This has been shown to be crucial in understanding... evolution.”

Even if one were to learn everything about the hominid-ape common ancestor, many of the most crucial questions about distinctively hominid evolution would remain unanswered. Although there is a fairly impressive record of human fossils during the period of 7–2 Ma, there is a lack of chimpanzee fossils at these early sites. It seems likely, therefore, that chimpanzee ancestors did not inhabit these fringe environments and were likely restricted to more wet, closed forest ecosystems—areas where fossils are less likely to be preserved. Some populations of chimpanzees moved into more mosaic, open habitats relatively recently, long after humans had moved into more arid environments. Modern chimpanzees do not live in habitats in which modern humans lived in the past or are found today; the historical geographic range of chimpanzees is quite restricted, probably more restricted than even that of early humans before leaving Africa. In our opinion, the best primate models to use as a basis for extrapolation about behavioral characteristics of our earliest ancestors are modern primate species living in similar edge habitats.

23.6 The Macaque model

Macaques can be extremely good colonizers of edge habitats. The macaque genus spread throughout Asia before humans reached that continent (Delson 1980). By the time H. erectus arrived in Asia 1.8 Ma, most hominids were no longer edge species (our more recent ancestors were exploiting more open habitats by this date), so hominids likely did not displace the macaques. True “weed” species, we propose that the macaques are excellent models for reconstructing how our early ancestors may have lived.

Long-tailed macaques (M. fascicularis) are small edge species that spend a good proportion of time both in the trees and on the ground. They are omnivorous and very versatile in their locomotion, although mainly quadrupedal. The most widespread of any Southeast Asian monkey, they occur from Burma through Malaysia and Thailand to Vietnam, while offshore populations are found on Java, Borneo, and numerous smaller islands as far east as the Philippines.
and Timor. Throughout this area, broadleaf evergreen and other forest types are interspersed with secondary and disturbed habitats, and it is the latter that long-tailed macaques prefer. Virtually all of the studies of this species make note that they are most commonly found in secondary forest habitat, preferably near water (Southwick and Cadigan 1972; Kurland 1973; Rijksen 1978; Crockett and Wilson 1980; Wheatley 1980; Rodman 1981; Sussman and Tattersall 1986). The success of the long-tailed macaque throughout its extensive Asian distribution is widely credited to its being an “adaptable opportunist” (Mackinnon and Mackinnon 1980 p 187). Researchers emphasize that these monkeys are extremely adaptable and able to flourish in highly disturbed land.

There is a sizable long-tailed macaque population on the island of Mauritius in the southwest Indian Ocean (Sussman and Tattersall 1981, 1986). Although the original transport of the species from Asia to Mauritius is undocumented, it is likely that they were on board ships when the Portuguese first reached the island and were inadvertently or purposely introduced to this ecosystem. Cited from the first studies as an assertive colonizer of new habitat, the small number of original immigrants has increased to 40,000 animals—successfully illustrating a facility to colonize disturbed and varied habitats (Sussman and Tattersall 1986).

Long-tailed macaques are slender, active monkeys; average weight is 4–5 kg for females and 6–7.75 kg for males (Jamieson 1998). Their society is organized around matrilineal hierarchies. One or two dominant males are visible within the group, as well as some lower-ranking adult males, plus the adolescent and subadult male offspring of the females. At sexual maturity, males migrate to a new group. Female offspring are philopatric, mating with unrelated males who join their troop (Jamieson 1998).

Many of the Man the Hunter models ignore or minimize the role of females in human evolution. In many terrestrial and edge-adapted nonhuman primate species, such as long-tailed macaques and baboons, females are the core of society, remaining with their female relatives from birth to old age. Each matriline consists of several generations and can be placed in a dominance hierarchy. Highly social females inherit the rank of their mothers, so the troop organization remains relatively stable over time.

Man the Hunter scenarios typically depict male hunters as leaders, innovators, tool makers, and tool users. Since these aspects of gender specificity may never be revealed in the fossil record, we believe it may be justifiable to construct theories based on our primate relatives. When Japanese macaques were first studied in the wild, it was the young females who started innovative behaviors such as new ways of processing food (Kawai 1958). Chimpanzee tools are made mainly by females and used mainly in gathering activities such as nut-cracking and termite-fishing; it is also the female chimpanzees who teach the
next generation how to use these tools (Boesch and Boesch-Acherman 2000; Sanz 2004). Furthermore, in most primates, females are the repositories of group knowledge concerning home ranges and scarce resources. Group knowledge and traditions are passed on from mother to offspring, and stability of the group, both in the present and over time, is often accomplished through female associations (Zihlman 1997).

In all primate species, males, females, or both sexes migrate between groups when they reach sexual maturity, thus maintaining genetic diversity within the population. In most edge species, such as long-tailed macaques, at sexual maturity males will leave the troop they were born into and seek an entirely new group of females who inhabit a different home range. Males in this kind of society will continue to migrate between groups throughout their lifetime. This situation portends a constant need for males to develop new social networks; often forming “friendships” with females in a new group as an entry strategy (Altmann 1980; Strum 2001).

In most primates adapted to edge environments, it is the males who migrate. However, in the closest genetic relatives of humans, the gorillas, chimpanzees, and bonobos, females normally change groups when they mature. This appears to be a phylogenetically conservative characteristic among hominoids which makes it possible that among our earliest ancestors females, not males, migrated between groups. However, most modern human foragers are multilocal with individuals residing with their maternal relatives at times and with their paternal relatives at other times, or sometimes with neither (Marlowe 2005).

The ability of edge species to exploit a wide variety of environments is accompanied by a substantial flexibility of behavior. Long-tailed macaques appear to be primarily arboreal where suitable vegetation exists but will descend to the ground along riverbanks, seashores, and in open areas. In some portions of their recently colonized range, such as Mauritius, they are highly terrestrial (Sussman and Tattersall 1981). They are eclectic omnivores with a distinct preference for fruit, but the variety of habitat they exploit is reflected in a wide selection of food items besides fruit, such as leaves, grasses, seeds, flowers, buds, shoots, mushrooms, water plants, gum, sap, bark, insects, snails, shellfish, bird eggs, and small vertebrates (Sussman and Tattersall 1981; Yeager 1996; Shaffer and Sussman 2005). Human-disturbed habitat or proximity to human settlements is not avoided; rather, they tend to live in proximity to humans throughout their range, which results in crop raiding of sugarcane, rice, cassava, and taro fields.

Long-tailed macaques live in large multimale, multifemale groups of up to 80 individuals, although in some areas groups are much smaller. They show distinct flexibility in structure; the large basic social unit tends to split up into smaller subgroups for daytime-foraging activities (Jamieson 1998; Sussman personal
observation). Subgroups may be all males but most often consist of adult males accompanying females and their young offspring. The number and size of subgroups tends to vary with the season and resource availability (Jamieson 1998). The entire troop reforms each evening and returns to the same sleeping site each night, usually on the edge of a water source. Because of their unique behavior of returning to a home base each night, long-tailed macaques could be labeled a “refuging” species.

### 23.7 Fossils and living primates

Looking at the fossil evidence, it is apparent that human ancestors, living between 7 and 2.5 Ma, were intermediate-sized primates, not smaller than male baboons or larger than female chimpanzees. Given their relative brain size, human ancestors were at least as clever as the great apes of today. They had diverse locomotor abilities, exploiting both terrestrial and arboreal habitats. They used climbing and suspensory postures when traveling in the trees and were bipedal when on the ground. We believe that their bipedalism was a preadaptation, but walking on two feet freed the arms and hands and proved to be advantageous in a number of ways.

Given their relatively small size and small canines, there is no reason to think that our early ancestors were any less vulnerable to predation than are modern monkeys—some of which have yearly predation rates generally comparable to gazelles, antelopes, or deer living in similar environments (Hart 2000). Edge species can be highly vulnerable to predation and because of this usually live in relatively large social groups with many adult males and females; adult males often serve as sentinels and provide protection against predators (Hart 2000). Because a primate group with only 1 male and 10 females can have the same reproductive output as a group with 10 males and 10 females, often the male acts as first line of defense; if a male is lost there are other males to take his place. If a sexually mature female is taken by a predator, then she and all her potential offspring (and living-dependent infants) are lost.

We propose that our early human ancestors may have lived in multimale, multifemale groups of variable size that had the capacity of splitting up, depending on the availability of food, and reforming each evening at home base refuges. It seems highly likely that females carried the young, conducted most of the socialization of the young, were repositories of group knowledge, had cognitive maps of the home range and its resources, were the center of society and the core of group stability, and spread innovations, techniques, and knowledge through the group and on to the next generation. However, certain facts, such as the exact size of the groups and subgroups, whether males or females migrated from their natal group, the internal structure of the group (whether matrilineal or formed...
along male kinship lines), would be impossible to determine accurately. Indica-
tions of these social parameters cannot be found in the fossil record and are quite
variable even in closely related living primates.

In sum, the best archetype of early humans may be a multimale, group-
living, mid-sized, omnivorous, quite vulnerable creature living in an edge habitat
near a large water source. These primates may well have been a refuging species
returning to the same well-protected sleeping site each night. Most modern
foragers are considered central place foragers, focusing their activities around a
central place, as are many birds, social carnivores, and primates (Marlowe 2005).
This early hominid was adept at using both the trees and the ground, but when
it exploited the terrestrial niche, it had upright posture and was bipedal. It
depended mainly on fruit, including both soft fruits and some that were quite
brittle or hard, but also ate herbs, grasses, and seeds, and gritty foods such as
roots, rhizomes, and tubers. A very small proportion of its diet was made up of
animal protein, mainly social insects (ants and termites) and, occasionally, small
vertebrates captured opportunistically. Our earliest ancestors did not regularly
hunt for meat and could neither process it dentally nor physiologically.

Like all other primates, and especially ground-living and edge species, these
eyearl humans were very vulnerable to predators and this trait did not diminish
greatly over time (Hart and Sussman 2005). Fossil evidence to this effect exists
from South Africa, from the Zhoukoudian cave in China, skulls uncovered at
Dmanisi in the Republic of Georgia, and a new fossil from Olorgesailie, Kenya
(Brain 1981; Berger and Clarke 1995; Gabunia et al. 2000; Boaz and Ciochon
2001; Small 2005).

23.8 Man the Hunted

Given that the earliest hominid ancestors were medium-sized primates who did
not have the morphological characteristics to fight off the many predators that
lived then—and given that they lived in edge environments which incorporate
open areas and wooded forests near rivers—then, like other primates, they were
vulnerable to predation. Because of this, we hypothesize that rates of predation
were just as high in our early ancestors as they are in modern species of primates
and our origins are those of a hunted species (Hart and Sussman 2005).

Protection from predation is one of the most important aspects of group-
living, and we believe this was true of our earliest ancestors. Based on the long-
tailed macaque model, it is probable that social groups of early hominids were
organized in a way that allowed efficient exploitation of a highly variable
and changing environment and also protected its members of the group from
predators. If the human lineage started out as Man the Hunted, we propose a
number of strategies for protection from predators based on the behavior and social organization we observe in long-tailed macaques. These include:

- Relatively large groups composed of 25–75 individuals: since safety lies in numbers, one of the main reasons all diurnal primates live in groups is predator protection which provides more eyes and ears alert to the presence of predators as a first line of defense. In his research on modern human foragers, Marlowe (2005) found that the median group size is 30 individuals.

- Versatile locomotion that exploits both arboreal and terrestrial milieus: the major advantage of agility in the use of diverse habitats is safety in trees and dense underbrush. An added advantage of upright posture is the ability to scan for predators.

- Flexible social organization: for example, gathering scarce resources in small groups but reuniting as a larger group when predation requires strength in numbers allows small groups to quickly disperse and hide while large groups can mob and intimidate predators. Again, modern human foragers fit this pattern of flexibility (Marlowe 2005).

- Multimale social structure: this demographic feature provides more male protection when traveling through open areas and when the group settles in evening or midday. When large groups break into subgroups, females and young are accompanied by one or more large males.

- Males as sentinels: males are usually larger in edge species. Upright posture adds to the appearance of large size and allows for better vigilance, as well as waving arms, brandishing sticks, and throwing stones. Being more expendable than females, males mob or attack predators.

- Careful selection of sleeping sites: refuging species bring the whole group together at night in a safe area. Staying in very dense vegetation during daytime rest periods is essential. Males stay on high alert during these inactive periods and when the group is on the move.

- Staying one step ahead of predators: intelligence endows primates with the ability to monitor the environment, communicate with other group members, and implement effective antipredator defenses (Hart and Sussman 2005).

This reconstruction of the behavior and ecology of our earliest hominid ancestors reflects the pervasive influence of large, dangerous predators throughout human evolution. Many circumstances have been proposed as a catalyst for the evolution of the human species—competition for resources, intellectual capacity, male–male conflicts, and hunting. However, looking at our primate relatives and the fossil record, we believe that predation pressure was one of the most critical components in shaping the evolution of our earliest ancestors.
References


Allen J, Stanford C (no date) Evaluating models and theories of hominin evolution. (Unpublished manuscript, Stanford personal communication)


Dart R, Craig D (1959) Adventures with the missing link. Harper, New York

Darwin C (1874) The descent of man, revised edition. The Henneberry Company, Chicago


Kurland JA (1973) A natural history of *Kra* macaques (*Macaca fascicularis* Raffles, 1821) at the Kutai Reserve, Kalimantan, Timur, Indonesia. Primates 14: 245–262


24 Modeling the Past: The Paleoethnological Evidence

Paolo Biagi

Abstract

This chapter considers the earliest Paleolithic, Oldowan (Mode 1), and Acheulean (Mode 2) Cultures of the Old Continent and the traces left by the earliest hominids since their departure from Africa. According to the more recent archeological data, they seem to have followed two main dispersal routes across the Arabian Peninsula toward the Levant, to the north, and the Indian subcontinent, to the east. According to the discoveries currently underway at Dmanisi in the Caucasus, the first Paleolithic settlement of Europe is dated to some 1.75 Myr, which indicates that the first “out of Africa” took place at least slightly before this date. The data available for Western Europe show that the first Paleolithic sites can be attributed to a period around 1.0 Myr. The first well-defined “structural remains” so far discovered in Europe are those of Isernia La Pineta in southern Italy, where a semicircular artificial platform made of stone boulders and animal bones has been excavated. The first hand-thrown hunting weapons employed by Homo erectus come from the site of Schoeningen in north Germany, where the occurrence of wooden spears, more than 2-m long, has been recorded for the first time from a site attributed to some 0.37 Myr. At roughly, the same time H. erectus is supposed to have begun the domestication of fire. Although most of the archeological finds of these ages consist of chipped stone artifacts, indications of art demonstrations seem to be already present in the Acheulean of Africa and the Indian subcontinent.

24.1 Introduction

The aim of this chapter is to review the current evidence for the paleoethnology of the early hominids who inhabited the Old World from the time of their appearance up to the end of the Middle Pleistocene. Although the data presently available are not abundant, there is no doubt that they are of key importance for
the understanding of early hominid behavior and lifestyles. The evidence concerned is limited in most cases to stone tools and their contexts (Clark 1968 p 277), almost exclusively due to natural and environmental factors both physical and biological (Stiles 1998 p 134). Early stone tool assemblages are often associated with alluvial sedimentary processes (Isaac 1967) related to the geographic and geomorphologic location and distribution of the (sometimes ephemeral) sites that in many cases are limited to the stone tools themselves, and possibly to organogenic tools and the faunal remains derived from hunting and scavenging activities. Nevertheless, the excavations carried out during the last 50 years, and the study of the settlement structures and tool assemblages of the Early Paleolithic sites of the Old World, “have shown that it is quite possible to find sealed occupation sites that have suffered little or no natural disturbance before or after burial” (Clark 1968 p 276).

As far as the remains of material culture and their chronotypological characteristics are concerned, this chapter deals almost exclusively with Mode 1 (Oldowan) and Mode 2 (Acheulean) complexes (Clark 1994). Tools belonging to these two “modes” have been collected from a great number of sites, which are distributed between East Africa and the Indian subcontinent in the southeast, and Europe in the northwest (Otte 2000 p 111).

### 24.2 Out of Africa

Much has been published dealing with the spread of the first hominids and the radiometric dating(s) of the “out of Africa” dispersal(s) (Chauhan 2005). Nevertheless many questions are still unresolved, since “the triggers for the movement of humans out of Africa are not well known” (Bar-Yosef and Belfer-Cohen 2000 p 81). This state of affairs results from the absence or scarcity of reliable data from some of the key territories that hominids must undoubtedly have crossed to reach Eurasia (Petraglia 2003: Figure 12).

This is the case for Arabia, from which little information is currently available, especially from the southern portions of the peninsula, more precisely Yemen (Dhofar) and Oman. Effectively, the Early Paleolithic sites discovered in these countries come from a few, restricted areas where intensive surveys and excavations have been carried out in the last two decades (Amirkhanov 1991; Cremaschi and Negrino 2002; Whalen et al. 2002; Whalen and Fritz 2004). Even though many of them are represented by surface finds, the Soviet–Yemeni Archaeological Mission excavated thick sequences in some caves of southeast Yemen, close to the Dhofar border. This led to the discovery of stratified complexes, which Amirkhanov (1994 p 218) attributed to the pre-Acheulean (Oldowan: Mode 1)
and Acheulean (Mode 2) periods. In this context, the only tool bearing evident traces of use, from the lowermost layers of Al-Guza Cave in Yemen (Amirkhanov 1991 p 107), is of unique importance. This is the only pre-Acheulean worn chopper so far known from the entire south Arabian Peninsula.

Although the Early Paleolithic sites so far discovered in this region are few, south Arabia is claimed to represent one of the key routes followed by the first hominids once they started to move out of Africa reaching the central territories of the Indian subcontinent, not later than 1.0 Myr (Bar-Yosef and Belfer-Cohen 2000 p 82). A second route is said to have been followed “across the Sinai into western Asia . . . although this has not been adequately detailed to date” (Bar-Yosef 1994 p 237; Petraglia 2003 pp 168–169), where the oldest site known to date is located at Ubeidiya (Bar-Yosef 1995 p 250) (Figure 24.1).

The two radiometric dates so far available from Dmanisi (Figure 24.2) in the Georgian Caucasus (Gabunia et al. 1999) demonstrate that this dispersal took place not later than 1.8 Myr (Gabunia 2000 p 43). Nevertheless “le mouvement oriental paraît à la fois beaucoup plus complexe et, surtout, beaucoup plus ancien qu’en Europe” (Otte 2000 p 108). Fortunately, the number of discoveries of Lower Pleistocene sites from this continent is systematically increasing (de Lumley 1976; Agustí et al. 2000; Mussi 2001 p 20).Although the absolute age of some of these

Figure 24.1
The Arabian Peninsula with the indication of the most important Early Paleolithic sites (dots) and the potential main routes followed by hominids during their “out of Africa” dispersal(s) (arrows) (after Petraglia 2003: Figure 12)
sites is problematic (Santonja and Villa 1990 p 54), many are undoubtedly much older than supposed only a few years ago (Roebroeks and van Kolfschoten 1994 p 500).

24.3 Chipped stone assemblages

24.3.1 Bifaces and other tools

As pointed out by Gowlett (2005 p 51), “East Africa is the key territory for examining the Oldowan and early Acheulean,” in which the first “bifacial tools were created about 1.5 million years ago” (Porr 2005 p 68) by Homo ergaster, as a consequence of a complex series of behavioral, economic, and social factors whose complexity has been pointed out by Porr (2005 p 77). Until recently, however, they have been considered almost exclusively in the context of “artefacts as a functional form that varies sometimes according to raw material considerations and is manufactured with a recurrent technology within broader parameters” (Ashton and McNabb 1993 p 190). But the fact that the manufacture of such tools continued for some 1.25 Myr indicates their importance, most probably not only as cutting and/or scavenging weapons (Domínguez-Rodrigo 2002) but also as social indicators independent of their functional meaning(s). According to Draper (1985 p 7), “we could imagine a situation where an Early
Paleolithic hominid might have fabricated a portable cutting tool for scavenging remnant meat from carnivore kills” that “was produced because a Middle Pleistocene knapper . . . was disposed to work stone in a way that produced an object we call a handaxe” (Hopkinson and White 2005 p 21). The high variability (Sinclair and McNabb 2005 p 185), the typological and dimensional characteristics (Isaac 1977), and the “wide temporal and geographic distribution” (Wynn 1995 p 11) of these tools have been noted by many authors, but from different perspectives and with different aims (Bordes 1968 p 23; Camps 1979; Petraglia 1998 p 371; McNabb et al. 2004; Hopkinson and White 2005) (Figures 24.3 and 24.4).

Although the complexity involved in the production of the lithic artifacts has been openly questioned (Hassan 1988 p 281), and the analysis of manufacturing techniques and debitage dispersal across the earliest Paleolithic sites (Gowlett 2005; Petraglia et al. 2005) is still rarely applied by the field archeologists, a few interesting exceptions should be mentioned. Among these is the MNK chert factory site in the Olduvai Gorge (Tanzania), which is dated to some 1.6 Myr. Here chipped stone artifacts, obtained from both local and imported raw materials, show a complex sequence of activities carried out by “early man working a raw material chosen for its technological properties brought to a central locality from diverse sources” (Stiles et al. 1974). FxJi50, in north Kenya, is a site 1.5 Myr old that “consists of a patch of stone artefacts interspersed with broken-up fragments of bone” (Bunn et al. 1980 p 111), whose precise function is still difficult to define. The chipped stone assemblage, which is composed of flaked cobbles and flakes, partly obtained on the spot, “has proved to consist of several dense clusters of material that interconnect with each other” (Bunn et al. 1980 p 114). This is one of the earliest Paleolithic sites from which “the close association (of bones) with artefacts and the presence of butchering marks suggest that the toolmakers were the first accumulating agency” (Bunn et al. 1980 p 125). This picture is rather unusual, if we consider that “for most of the sites excavated and reported we do not have certain indications of any specific activities that characterize them, and in very few instances has localization of subsidiary tool kits within a floor even been claimed” (Isaac 1972 p 185) and that the interpretation of the variability of the spatial distribution pattern of the tools (Whallon 1973 p 117) within a site surface is often difficult (Keeley 1991 p 258).

### 24.3.2 Raw material, workshops, and quarries

When detailed recording methods have been applied, for instance, in the case of some localities excavated in the Indian subcontinent, they have revealed that characteristic tools, among them hand axes, cores, hammerstones, and different
Figure 24.3
Variation among lower Paleolithic biface assemblages of eastern Asia and south Asia. The dashed line represents the Movius line, the traditional demarcation between Mode 1 (Oldowan) and Mode 2 (Acheulean) industries (after Petraglia 1998: Figure 11.8)
dimensional classes of debitage flakes, systematically cluster in well-defined spots. This fact is useful in helping us to understand the development of the manufacturing areas within the site and the steps followed by the toolmakers during the production process (Hansen and Madsen 1983 p 51), especially when refitting methods are applied to the entire complex (Bergman et al. 1990 p 280). This is the case for the some Acheulean sites where different varieties of raw materials for tool production were available, including siliceous limestone (Isampur in India: Petraglia et al. 2005) and good quality chert from local outcrops (Rohri Hills in Sindh [Pakistan]: Biagi et al. 1996).

The evidence available from the second case shows that the waste products of large hand-axe-manufacturing workshops were scattered along the edges of circular sandy areas, which represent zones that were comprehensively cleared of limestone and chert boulders in Paleolithic times, before the manufacturing activities took place. For instance, the excavations carried out at Ziarat Pir Shaban 1 (Figure 24.4), one of the many Acheulean workshops discovered on the Rohri Hills that were exclusively devoted to the production of hand axes (Biagi et al.
1996) (Figure 24.6), has demonstrated that the perfect, finished bifaces were exclusively transported elsewhere, most probably to camps located in the adjacent Great Indian Desert that are at present buried beneath meters of sand inside thick, stabilized dunes (Misra and Rajaguru 1989). The maximum transfer distance is not known, due to the absence of any detailed research in the Thar Desert to the east of the hills, although the African parallels indicate transport between 15 and 100 km (Petraglia et al. 2005 p 208). A situation similar to that of the Rohri Hills is known at Ongar, near Hyderabad in lower Sindh (Pakistan), where Acheulean workshops were discovered in situ on the top of flat, limestone mesas (Figures 24.7 and 24.8). These deposits, very rich in seams of excellent chert, were exploited throughout the entire Paleolithic period, from the Acheulean onward (Biagi 2006).

As far as these two latter cases in Sindh are concerned, there is no doubt that the abundance of excellent, workable raw material played a fundamental role in attracting prehistoric populations at least since the Acheulean period (Biagi and Cremaschi 1988 p 425). The chert used by the earliest Paleolithic people was collected from large boulders or extracted from the top of the limestone terraces, although the accurate surveys carried out along the top of the mesas did not reveal any trace of Early Paleolithic mining activities.
Figure 24.6
Ziarat Pir Shabân on the Rohri Hills (Sindh, Pakistan): Acheulean hand-axe rough-outs on the surface of workshop ZPS1 (photograph by P. Biagi)

Figure 24.7
Ongar (Sindh, Pakistan): C-shaped Acheulean chert factory area (photograph by P. Biagi)
As far as we know, the first Paleolithic chert quarries were opened by the Acheulean populations in Upper Egypt much earlier than until recently supposed (Smolla 1987 p 129). According to Vermeersch et al. (1995 p 22), “a few kilometres south of the Dandara temple ... a hill was clearly subjected to chert extraction by Acheulian people,” given the presence of an extractive pit discovered during the excavation of a small trench in an area rich in Late Acheulean tools. In contrast, almost nothing is known of the Acheulean raw material procurement systems in this region, which yielded abundant traces of Middle and Upper Paleolithic flint-mining activities (Vermeersch et al. 1997 p 191).

24.4 Habitation and other structural remains

Early Paleolithic Mode 1 and 2 sites are often characterized by “concentrations of debris, ... which ... have usually been interpreted to be the result of various processual phenomena” (Stiles 1998 p 133). Only a few of them, of varied chronology, have provided us with complex archeological evidence.
In Africa, we know that most of the earliest settlements were located in environments close to lake shores or, more commonly, along (former) river courses (Isaac 1976: Figure 3.3) (Figure 24.9). They have been interpreted as sites that are inhabited during only one season. The 1.75 Myr old Mode 1 site of DK, in Lower Bed I of the Olduvai Gorge (Leakey 1971 p 24 and Figure 7) yielded evident traces of man-made features, the most important of which consists of a circular structure of lava blocks, some 4.5 m in diameter (Figure 24.10), which the excavator interpreted as resembling “temporary structures often made by present-day nomadic peoples who build a low stone wall round their dwellings to serve either as windbreak or as a base to support upright branches which are over and covered with either skin or grass” (Leakey 1971 p 24).

The excavations carried out at Gomboré I, another Mode 1 site located at Melka Konturé in Ethiopia, brought to light a 230 m² living floor composed of rounded pebbles, rich in stone tools and faunal remains with an empty space of some 10 m² in between. The settlement, which has been dated at some 1.6 Myr, yielded a “higher platform . . . that . . . could have been roughly adapted for a shelter made of branches and animal skins” (Chavaillon 2004 p 263). The research carried out at this site revealed the occurrence of “small stone circles aligned north-south in the eastern sector . . . whose . . . external diameter . . .
Figure 24.10
Olduvai Gorge, site DK (Tanzania): Plan of the stone circle and the remains of the occupation surface: Stone artifacts shown in black, bones in outline (after Leakey 1971: Figure 7)
varies from 20 to 40 cm,” which were interpreted as possible “wedging stones for pegs set in rather hard soil” (Chavaillon and Chavaillon 2004 p 448). The presence of almost identical features was recorded also from the Acheulean site of Garba XII in the same region of Ethiopia.

Among the Mode 2 sites, extremely interesting and ideally preserved remains were brought to light at Isernia La Pineta in Molise (southern Italy). This site, whose chronology is still rather controversial (Mussi 2001 p 44), although the new radiometric dates indicate that the site is some 0.60 Myr years old (Coltorti et al., 2005; 19), extends over an area of some 30,000 m². It yielded traces of four different occupation layers from which more than 10,000 lithic artifacts, chipped from different raw materials, including limestone and chert from diverse sources, were collected (Peretto 1994). The site was located along the shores of a lake-basin, later buried by fluvial sediments. The most interesting structural remains were discovered during the beginning of the excavations, when an accumulation of animal bones and stone tools was uncovered on an almost semicircular paleosurface that was very rich in remains of Bison skulls and horns and Rhinoceros cranial bones and was delimited by large, travertine boulders (Giusberti et al. 1983 p 100) (>Figures 24.11–24.13). These discoveries might help interpret the spatial variability and activities carried out within this settlement site (Bartram et al. 1991). Remarkable differences among the lithic assemblages have been noticed between the different excavated areas, both in the raw material employed for producing artifacts and in the typology and dimension of the stone tools (>Figure 24.14) (Peretto 1983 p 81).

In central Italy, an interesting Mode 2 site dated to slightly later than 0.5 Myr, and with an assemblage consisting of both elephant long bones and stone bifacial hand axes, has been excavated at Fontana Ranuccio (Biddittu et al. 1979). The presence of bone hand axes is unique to the area (Biddittu 1982), where they become increasingly more common at the slightly later Mode 2 sites, like Castel di Guido in Latium (Radmilli and Boschian 1996).

Moving westward, the importance of the remains of structures, brought to light by H. de Lumley (1966) at Terra Amata, near Nice, in Provence, is represented by a shallow, oval-shaped hut-floor attributed to a Mode 2 group of people who inhabited the region around 0.4 Myr. Apart from the exceptional discovery of an almost “intact” habitation structure, the site is important because it yielded the first evident traces of a hearth indicating the domestication of fire by Paleolithic man in Europe, although traces of fire have long been known from a few Lower Pleistocene sites in East Africa (Clark and Harris 1985; Perles 1977).

The site of Bilzingsleben, in Germany, is of extreme importance for the study of the behavior of Homo erectus. The remains of three circular hut foundations, 3–4 m in diameter, with entrances systematically facing southeast, and with
Figure 24.11
La Pineta (Isernia, southern Italy): A general view of the semicircular animal bones and material culture remains concentration surrounded by limestone boulders, discovered in 1980 (photograph by P. Biagi)

Figure 24.12
La Pineta (Isernia, southern Italy): Bison skull and long bone fragment from the main semicircular concentration discovered in 1980 (photograph by P. Biagi)
Figure 24.13
La Pineta (Isernia, southern Italy): plan of the concentration of figure 24.11: (A) travertine, (B) pebbles, (C) faunal remains, (D) limestone tools, (E) flint tools, (F) red lacquerings (after Giusberti et al. 1983)
workshop areas and fireplaces, have been discovered at this camp, dated to some 0.37 Myr (Figure 24.15). The importance of this site is indicated by the occurrence of the earliest so far known intentionally decorated bone objects that suggest, “non-utilitarian behaviours ... connected to reflexive thinking” (Mania and Mania 2005 p 110), as well as the indisputable traces of what is claimed to be a ritual paved area “with human skull fragments smashed in macerated condition” (Mania and Mania 2005 p 113). These discoveries demonstrate that “Homo erectus was therefore a human being that had a fully developed mind and culture, capable in creating his own socio-cultural environment with living structures, the use of fire and special activity areas” (Mania and Mania 2005 p 114). This also finds confirmation in the traces of Acheulean “art” both in Africa (Bednarik 2003) and in the Indian subcontinent (Bednarik 1990).

Gran Dolina at Atapuerca in Spain is an even earlier multilayered site, where some kind of ritual activity has been supposed to have taken place. The site yielded 150 human bone fragments, which have been attributed to four individuals, classified into the new form Homo antecessor. Some of the hominid
remains from Layer TD6, datable to at least 0.8 Myr, “show clear cut marks which have been interpreted as evidence of cannibalism” (Mosquera Martínez 1998 p 17).

Returning to Mediterranean France, this region is very rich in Lower Paleolithic sites, both open air and in caves. Among the latter, the internal deposits of Lazaret Cave (de Lumley 1969), a late Mode 2 Acheulean site attributed to some 0.12 Myr, yielded traces of a unique hut structure that has been reconstructed thanks to the occurrence of stone walls, fireplaces, and “masses of seaweeds possibly used as bedding for site occupants” (Mellars 1995 p 285). Although this site does not represent the earliest known evidence of cave structural remains in Eurasia, given the traces of much older man-made stonewalls in China (Fang et al. 2004: Figure 3) and Central Europe (Cyrek 2003: Figure 6), Lazaret is the

Figure 24.15
Bilzingsleben (Germany): Plan of the structuration of the Early Paleolithic camp: (a) limits of the excavated area, (b) geological fault lines, (c) shoreline, (d) sandy travertine sediments, (e) alluvial fan, (f) activity area at the lake shore, (g) outlines of living structures, (h) workshop areas, (i) special workshop area with traces of fire use, (J) circular paved area, (k) charcoal, (l) bone anvils, (m) stone with traces of heat, (n) bones with intentional markings, (o) linear arrangement of stones, (p) elephant tusk, (q) human skull fragments, (r) human tooth (after Mania and Mania 2005: Figure 7.1)
only one from which a detailed reconstruction of the events that took place inside the cave in Late Acheulean times have so far been possible.

24.5 Hunting weapons

Although, as mentioned earlier, the excavations carried out at Terra Amata in the 1960s had already revealed the presence of one single fireplace, the almost contemporary hunting site of Schöningen, in North Germany, yielded not only the remains of four hearths, one of which is some 1 m in diameter, but even a charred wooden stick, which might “have functioned as a firehook to feed the fire as well as a spit to roast, and also smoke, strips or pieces of meat” (Thieme 2005 p 127). This site is extremely important because of the occurrence of both the hunting weapons and the other wooden tools brought to light since 1994, which have radically revolutionized our view of the hunting methods and strategies followed by these hominids. The widely accepted view that *H. erectus* was unable to conceive and construct throwing weapons is contradicted by the discovery of sophisticated spears, even longer than 2 m, which suggest a long tradition in wood shaping and weapon craftsmanship showing that, in contrast to what was previously supposed, this species had already acquired that complex “sequence pattern of behavioural complexes” (Laughlin 1968 p 305) commonly labeled hunting, which represent “a way of life . . . that . . . has dominated the course of human evolution for hundreds of thousands of years” (Washburn and Lancaster 1968 p 293). More precisely “*Homo erectus* in the Middle Pleistocene was fully capable of organising, coordinating and successfully executing the hunting of big game animals in a group using long-distance weapons” (Thieme 2005 p 127). Although the Schöningen specimens are not the only wooden pointed tools so far recovered from an Early Paleolithic site in Europe, they undoubtedly represent the best preserved specimens discovered within a horse-hunting camp, a surface of some 3,500 m² of which has already been excavated.

Furthermore it is important to point out that already in the 1980s, Isaac (1984 p 17) had considered the use of throwing weapons by early hominids when he wrote “if the Lower Pleistocene tool-making hominids were hunting with equipment, they must have been using spears without stone tips (i.e. pointed staves or horns on staves), clubs, and, perhaps most important of all, thrown sticks and stones,” given that “none of the flaked stone artefacts can plausibly be regarded as ‘weapons’” (!). In effect it has been widely demonstrated that stone hand axes and cleavers are excellent butchering tools, but not hunting weapons, and, in particular, that “the sinuous retouched edge of a hand-axe retains its meat-cutting efficiency longer than a plain flake edge” (Isaac 1984 p 15).
24.6 Any ethnographic parallel?

Apart from the factors mentioned in the introduction, there are many others that make remains of early structures difficult to interpret. Among these are (1) the impossibility of “detailed” radiometric dating of the events that took place at short-term habitation sites and (2) the difficulty of proving the supposed contemporaneousness of the structural remains within an apparently “homogeneous” area. This is true even though it is widely assumed that “in inspecting the contents of a single structure, we can be fairly confident that the associated assemblage was all in use at one time, if not made at the same time” (Deetz 1968 p 283). Besides the two above-mentioned factors, there are two others of major importance regarding (1) the complete excavation of an occupation unit, an enterprise that has been successfully undertaken only on very few occasions (Clark 1968 p 277) and (2) the functional nature of the (seasonal) site itself (Hehmsoth-Le Mouël 1999 p 81).

With the exception of a limited number of cases reported by Clark for East Africa, and a few others which have been described in the preceding chapters, most sites are characterized by more or less dense concentrations of stone artifacts, differently disposed according to the activities performed (Stevenson 1991 p 280), reflecting “a complex system of extraction, manufacture, transport, use, resharpening, re-use, renewed transport and eventual discard” (Isaac 1986: Figure 15.6). Often, these have been subjected to a certain degree of weathering or represent a (complicated) sequence of depositional events that took place over a period of millennia. Isaac (1968 p 255) classified such concentrations in three main categories according to the vertical and/or horizontal diffusion of the stone tools. The first two of these “represent sporadic, intermittent occupations of great duration,” while the third “can probably be interpreted as fairly stable ‘home base.’”

Finally, ethnographic analogies are sometimes uncritically accepted by both archeologists and anthropologists, who often believe “that modern representatives of past stages of cultural development exist” (Freeman 1968 p 263), even though “any consideration of the implications for archeological interpretation of new ethnographic data . . . requires an examination of the general relationships between ethnographic observations and archeological reasoning” (Binford 1968 p 268).

References


Archaeological investigations of Lower and Middle Palaeolithic landscapes, locales and artefacts. Routledge, London New York, pp 13–28


Hyman, Boston Sidney Washington, pp 343–390


25 Modeling the Past: The Linguistic Approach

Bernard Comrie

Abstract

First, traditional methods of comparative-historical linguistics are discussed as a means of reconstructing aspects of the past through linguistic means, with particular regard to the comparative method, the Neogrammarian hypothesis of the regularity of sound change, the construction of family trees of genealogically related languages, and linguistic paleontology as a means of reconstructing the physical and cultural environment in which the ancestor language was spoken. More detailed consideration is given to the reliability of different kinds of evidence in the application of these methods, especially problems arising from language contact and resulting horizontal rather than vertical transmission of innovations, and from the gradual splitting of an ancestor language into its descendants and the proposed wave theory model of this process. Emphasis is placed on the need for detailed and painstaking investigation in order to apply these methods. A critical discussion is then given of some more recently proposed methods, in particular mass comparison (with its lack of emphasis on the need to rely on the regularity of sound change), glottochronology (with its assumption of a constant rate of replacement for basic vocabulary), and typological comparison.

25.1 Introduction

Modeling and reconstructing the past is a traditional task in linguistics that goes back at least to the early nineteenth century, and which by the late nineteenth century had led to an impressive series of results, perhaps most notably the reliable identification of the Indo-European language family. This language family subsumes most of the languages spoken in Europe, northern South Asia, and some intermediate areas. The Indo-European hypothesis postulates that all these languages derive from a common ancestor by means of changes that can, to a large extent, be reconstructed even in the absence of direct attestation in historical documents. The methodology has also been applied to other languages, establishing other language families and a range of general principles governing
the ways in which languages can change. The subdiscipline of historical linguistics covers all aspects of language change. Historical-comparative linguistics is concerned more narrowly with those subparts of historical linguistics that are relevant to the grouping of languages into language families. The aim of this chapter is to describe and illustrate the basic principles that underlie historical-comparative linguistics as a tool for modeling the past, including both widely accepted methods and, with due caveats, some other methods that have been proposed more recently. A subsidiary aim is to try and explain why many linguists have been reluctant to employ some of these more recently suggested methods, often borrowed from other disciplines like archeology and genetics, while not forgetting that to understand motives is not necessarily to endorse them. Given the restricted space available, the account is necessarily illustrative rather than exhaustively argued. There are, incidentally, several standard introductions to (comparative-)historical linguistics, including Campbell (1998), which provide more detailed justification or references thereto.

25.2 Comparative method

25.2.1 Basic principles

The standard, most widely accepted method for establishing that languages are related via a common ancestor and for reconstructing that common ancestor and the changes separating the attested languages from their common ancestor is the comparative method. It will be useful to introduce the comparative method in several steps, some of which are provisional approximations that will be elaborated in later discussion. When two or more languages are claimed to descend from a common ancestor, it is usual in linguistics to say that these languages are “genetically related”; to avoid any possible confusion between this use of the term “genetic” and its use as the adjective corresponding to “gene,” “genetics” in biology, I will prefer a nontraditional terminology and say that such languages are “genealogically related.”

The comparative method resides in seeking traces of the putative common ancestor in the attested descendant languages. As a first approximation—which will require several refinements—we may say that if we find similarities in two or more languages that we are investigating, then we can, at least provisionally, attribute those similarities to their common ancestor. If we find more similarities between languages A and B than between either of these and C, then we may hypothesize further that A and B have a common ancestor—let us call it D—that is not shared by C, while D and C in turn share a common ancestor, which we may call E. This leads to a hierarchical structure as in Figure 25.1.
Historically, this can be interpreted as follows. At one time there was a single common ancestor, E. At some point in time, E splits into two descendant languages, D and C, because of different innovations taking place on each of the branches E–D and E–C. Subsequently, D in turn splits into two descendant languages, A and B, each in turn characterized by the different innovations of the branches D–A and D–B. The kind of hierarchical structure that results is known as the family tree of that particular language family.

In interpreting a linguistic family tree, it is important to bear several factors in mind, some of which will lead to refinements of the account given so far of the comparative method. One of the most important is that the branches of the diagram are defined in terms of common innovations. The justification for the node D, for instance, is not just that A and B share certain features that are attributable to their common ancestor D, but more specifically that these features are innovations that must have characterized D relative to E. A concrete example will make this clear. The Germanic languages (English, German, Swedish, etc.) are distinguished from other Indo-European languages by having $f$ where the other languages have $p$, compare English *father*, *fish*, *foot* with their Latin equivalents, respectively, *pater*, *piscis*, *pes*. Given that the languages that have $p$ in such words do not belong, on independent grounds that cannot be argued here, to a single branch, we can conclude that the $p$ in these words represents the initial state, and that the Germanic languages have undergone a shift from $p$ to $f$. (This is, incidentally, part of a radical change in the consonant system of Germanic languages relative to other Indo-European languages, known traditionally as Grimm’s Law.) Going back to the family tree discussed earlier and interpreting it in terms of the current discussion, we might say that A and B are English and German respectively, with D their common ancestor Proto-Germanic, while C is Latin, and E the common ancestor Proto-Indo-European. (It is usual to label the common ancestor of a group of languages with the prefix “Proto-.” Of course, there are many other Germanic languages, and many other Indo-European branches and languages, which are not represented in this partial diagram.)

The importance of common innovations, rather than just similarities, cannot be overestimated. In the example just discussed, the Germanic languages form a branch of the Indo-European family because they share a common innovation ($p > f$). The Indo-European languages that have $p$ do not form a
single branch, rather they are the residue of languages that have not undergone the shift that characterizes Germanic. This is, incidentally, at the root of many linguists’ discomfort with methods imported from other disciplines that rely simply on different states, e.g., with p and f as different states in the example just discussed, without considering which of the states is original and which is an innovation. In some cases, of course, it may be difficult or impossible to decide which of two states is the innovation, and in such cases it is correspondingly difficult or impossible to use that feature as the basis for constructing a tree.

25.2.2 Regularity of sound change

It will be noted that in the example just discussed of Grimm’s Law, it is not simply the case that we have a single instance of the correspondence p–f, in a single word, rather this same correspondence p–f is found in a range of words. In fact, we can take this principle further and say that, ideally, the same correspondence should hold in all words. This is the principle of the regularity of sound change, also known as the Neogrammariann hypothesis, one of the cornerstones of comparative-historical linguistics. More specifically, the principle of the regularity of sound change says that a given sound (in our example, p) will always change in the same way (in our example, to f) in the same environment. The addition of the condition “in the same environment” is crucial, because the same sound can develop differently in different environments. In the development to the Germanic languages, for instance, original p did not shift to f after s, so that we find correspondences like English spew, Latin spuere “to spit,” where Germanic p, not f, corresponds to p in other Indo-European languages. Although some linguists have recently questioned the value of the principle of the regularity of sound change, most notably those associated with the Greenberg approach (see 25.4.1), its value is nonetheless recognized by most scholars working in the area.

The principle of the regularity of sound change actually has a number of important corollaries for the application of the comparative method. Most importantly, what we are examining are not literally similarities across languages, but rather regular correspondences. First, items might be similar between two languages, but if this is not part of a regular pattern then they do not constitute a regular correspondence that can form the basis of establishing a family tree. Perhaps not surprisingly, words can sometimes resemble one another purely by chance. But some examples can be surprising if one is caught unaware. Take, for instance, English much and Spanish mucho, both with the same meaning and almost the same form. Even though English and Spanish are in fact genealogically related languages, though belonging to different branches of the Indo-European language family—Germanic and Italic (Romance), respectively—these two words
do not derive from a common ancestor, and as we trace their history back through time they look less and less like one another. English *much* is ultimately related etymologically to Latin *magnus* and Greek *mégas* “big,” while Spanish *mucho*, which derives from Latin *multum* “much,” is ultimately related to Greek *mála* “very.”

Conversely, words across languages can sometimes be related by means of regular correspondences even though the words themselves do not look particularly similar. It would be hard to find two words that look less like one another than the Armenian word for “two,” erku, and its Russian equivalent *dva*. In fact, however, the initial *erk-* of the Armenian form goes back quite regularly to Proto-Indo-European *dw-*, this same correspondence being found in a number of other words. (In historical linguistics, it is usual to prefix an asterisk to forms that are reconstructed, not attested. Armenian forms an independent branch of the Indo-European family, while Russian belongs to the Slavic, or Slavonic, branch. Note that English *two* is also related etymologically to the same Proto-Indo-European root. The examples in the last two paragraphs are taken from Ringe (1999), a useful reference in the present context in that it is intended for an interdisciplinary audience.) It will be observed that the establishment of regular correspondences, rather than superficial similarities, can be a painstaking process, one of the reasons why linguists are often suspicious of methods (such as those to be discussed in Section 25.4) that promise shortcuts.

There are, however, some exceptions to the principle of the regularity of sound change that seem just to have to be recognized as such. For instance, the Modern English pronoun *she* is almost certainly derived etymologically from Old English (Anglo-Saxon) *sēo* (with a number of attested variants in the vowel of the Old English word, but not the initial consonant), although the initial consonant of the Old English word would regularly have given Modern English initial *s-*, not *sh-. While there is a rather plausible series of idiosyncratic changes that will get from Old English *sēo* to Modern English *she*, this does involve going outside the strict limits of regular sound changes. This means that there is a certain amount of “noise” in historical reconstructions. While such noise is relatively small across brief time periods—Ringe (1999) estimates that between Old English and Modern English only 2% of sounds have changed irregularly—the accumulation of such exceptions across longer periods of time does mean that a limit will eventually be reached beyond which it is impossible to identify regular sound change because of the number of exceptions.

In some cases, we can provide an explanation for an irregular development that goes beyond the mere consideration of sounds. Consider, for example, the development into Modern English of the Old English verbs *cniawan* “to know,” past tense *cnēow*, and *māwan* “to mow,” past tense *mēow*. The forms of “know” regularly give Modern English *know*, past tense *knew*. By regular sound change,
the forms of “mow” would be *mow*, past tense *mew*, but in fact the past tense is *mowed*. What has happened here is that the originally irregular past tense of *mow* has been replaced by the regular formation in *-ed*, by a process that is known as analogy. Incidentally, *know* in Modern English also has an irregular past participle, *known*; in the case of *mow*, analogy has been partial, since both the expected irregular form *mown* and the analogical regular form *mowed* are found. Thus, although *mowed* is not the regular development of Old English *mēow*, we can explain how the Modern English form came about by appeal to other factors, admittedly such as do not apply regularly: the irregular past tense *knew* has not been regularized analogically.

### 25.2.3 Bases of cross-linguistic similarity

Chance is one way, in addition to common ancestry, in which words in two languages can show up as similar to one another. Another basis for similarity can be universal features of language, either in the sense of absolute universals, i.e., features that are common to all languages, or universal tendencies, i.e., features that have a widespread distribution independent of any particular historical event. If a feature is an absolute universal, then it of course plays no role in establishing the hierarchical structure of language relatedness, since all languages will by definition have the same value for this feature. Widespread ahistorical features are potentially more problematic, even if their number is probably rather restricted. One widespread cross-linguistic similarity that may well find its explanation here is the occurrence of words like, or at least beginning with, the sequence *ma* and meaning “mother”; compare, for instance, Latin *mater*, Mandarin Chinese *mā* and Haruai *mam*—Haruai is an indigenous language of highland New Guinea. While this might be historically determined, there is an alternative explanation, namely that this is one of the first sound sequences usually uttered by human infants, and that its interpretation to mean “mother” has been carried out independently by parents across the world. Of course, if this is the case, then there might be other interpretations that could be assigned to this sequence, and we find languages like Georgian, the national language of the Republic of Georgia, in which *mama* means not “mother” but “father.” (The Georgian for “mother” is, incidentally, *deda*.)

However, by far the most troubling factor in identifying common ancestry is the possibility of borrowing, whereby one language takes over a word from another language. (Borrowing can also apply to linguistic units other than words, e.g., syntactic constructions, but here it is primarily with words that we need to concern ourselves.) In the example, discussed earlier, of the regular
correspondence between Germanic \( f \) and \( p \) in other Indo-European languages, the careful reader might have observed that while this works well for English \textit{father}–Latin \textit{pater}, there seems to be a problem with English \textit{paternal}, which begins with a \( p \), instead of the expected \( f \) (cf. \textit{fatherly}). The explanation is that English \textit{paternal} is not transmitted vertically from Proto-Indo-European via Proto-Germanic but is rather transmitted horizontally by being borrowed from Latin, more accurately from Medieval Latin \textit{paternalis} (perhaps via French \textit{paternel}). In this particular case, we can trace fairly accurately the historical developments in English texts over the centuries so that the borrowing process that led to the Latin loan word \textit{paternal} in English can be readily identified. There are other criteria that can also be used. In particular, the regularity of sound change, more specifically the regular \( p–f \) correspondence between Latin and English, allows us at least to suspect that in English \textit{paternal} we are not dealing with a true etymological cognate of the Latin form, but rather with a loan, given the unexpected \( p–p \) correspondence without a preceding \( s \).

There can, however, be more difficult cases. In particular, if one language borrows a substantial chunk of vocabulary from another over a limited period of time, then the result may actually lead to a set of regular correspondences that would be misleading if one did not know the actual course of events. For instance, in the prehistoric period Finnish borrowed a number of words from Russian. Russian has an opposition between such consonants as \( p \) and \( b \) which is missing in Finnish but was interpreted by Finnish speakers in terms of their own opposition between \( pp \) and \( p \) between vowels, which leads to regular correspondences intervocalically between Russian and Finnish of \( p–pp, b–p \), as in Finnish \textit{pappi} < Russian \textit{popu˘} “priest,” Finnish \textit{liepä} < Russian \textit{xle˘bu˘} “bread.” (In the word-initial environment, both Russian \( b \) and \( p \) appear as \( p \) in Finnish loans of this period; Finnish does not allow word-initial \( pp \).) It should be noted that much of the current controversy between proponents and opponents of larger language families, such as Altaic (incorporating at least the Turkic, Mongolic, and Tungusic languages, perhaps also Korean and Japanese), hinges on whether similarities are attributed to common ancestry or borrowing.

Various attempts have been made to try and establish particular sets of vocabulary that are immune or, more probably, less susceptible to borrowing. For instance, it seems to be generally true that more basic vocabulary (such as \textit{arm} versus \textit{ulna}) is less likely to be borrowed, even if there are no concepts that can be characterized absolutely as unborrowable. Hitherto, the range of languages that has been considered in the formulation of such generalizations has been relatively restricted, in particular because until recently few language families have been investigated in sufficient detail to provide reliable etymological material. Recent advances in the etymological investigation of a wider range of
language families mean that it is now possible to test such hypotheses more reliably, and a project of just this nature, the Loanword Typology Project, is currently underway at the Max Planck Institute for Evolutionary Anthropology; see http://www.eva.mpg.de/lingua/files/lwt.html.

More generally, one can say that the problem in applying the comparative method is to find similarities or correspondences that are more plausibly attributed to common ancestry than to any other cause. One reason for choosing words as the basis of the method is that, a priori, we would not expect there to be any similarity between the shape of the word expressing a particular concept in one language and the shape of the corresponding word in another language, given the general principle of the arbitrariness (or conventionality) of the linguistic sign—there is no reason why a dog should be called dog rather than inu (Japanese) or mbwa (Swahili). A corollary of this is that, in applying the comparative method, we need to steer clear of phenomena that are likely to occur independently on different occasions. For instance, a sound change like $p > f$, which characterizes the Germanic languages, is relatively rare cross-linguistically, especially when taken in conjunction with the other sound changes that constitute Grimm's Law, and is therefore a good diagnostic as a common innovation. Changes that more frequently occur independently are less good diagnostics. For instance, in the history of both English and German a change has taken place that shifted the long vowel $\ddot{u}$ to the diphthong $au$, as in Modern English house, Modern German Haus, both pronounced /haus/, compare Old English and Old High German $h\ddot{u}s$. These two changes occurred independently, long after both English and German had split off from Proto-Germanic. Some dialects of both languages, such as Scots English and Swiss German, retain the undiphthongized long vowel. The “common innovation” of the diphthongization of $\ddot{u}$ to $au$ is therefore not usable as a criterion for grouping English and German together. In other words, the kinds of sound changes that are most useful are those that are least usual, since these are least likely to occur independently in two different historical events.

In the examples considered so far, the words that have been compared have usually been translation equivalents, i.e., in more informal terms have the same meaning. One example did, however, involve a slight change in meaning that was glossed over at the time, namely the pair English spew–Latin spuere, where the meanings are not quite identical, the English verb meaning “vomit,” the Latin verb “spit.” In this particular case, we happen to know from the historical record that the meaning of the English verb has undergone semantic change over the history of the English language, since in Old English it could also mean “spit.” Where languages are known to be related through correspondences that do not involve semantic shifts, then words that are similar in meaning and that show regular sound correspondences can usually be assumed to be good cognates, as in
the case of the English–Latin pair just mentioned. But the principles of semantic change are much less well worked out than those of sound change, and it is even possible that there may be no principles that can be set out with comparable rigor. To take one striking example, in less than 1,000 years the English adjective *nice* has shifted in meaning from “ignorant” to “good.” The historical development is well attested in English texts over the centuries, but this does not license us to claim an etymological relation between any phonetically similar pair of words meaning “ignorant” in one language and “good” in another. Many attempts to establish wide-ranging language families have been criticized for laxity in the treatment of the semantics of the items compared.

One final comment is in order. It is, of course, quite possible that a particular language, or even a branch of a language family, will have lost a word that existed in the protolanguage. It follows from this that words do not have to be found in all descendant languages, or even in all descendant branches in order to be projected back to the reconstructed protolanguage. Of course, if a word is found in only one language or branch, we have no evidence for projecting it back beyond that language or branch (even though we cannot exclude the possibility that it might go back further and have been lost, perhaps independently, in other languages and branches). How widespread a word needs to be across a language family in order to be reconstructed to the protolanguage is a complex question, though one that can be answered at least partially at least for some language families. In the case of Indo-European, for instance, a word found in a branch that is held to have separated earlier from the rest of the family (such as Anatolian, a now extinct branch whose best known constituent language is Hittite) and in the rest of the family can probably be reconstructed to the protolanguage. By contrast, a language found in even a number of the western branches might reflect a local innovation rather than an inheritance from the protolanguage.

### 25.2.4 Mixed languages

The discussion so far has assumed that any language can be assigned unequivocally to one language family or another. Parts of a language may be borrowed from another language belonging to a different family (or branch of the same family), but the very notion of borrowing assumes that there is a core to the language that remains genealogically inherited from its ancestor. An obvious question to arise is whether there can be exceptions to this generalization, i.e., languages that have taken substantial parts of their overall structure from different sources to such an extent that it is impossible or arbitrary to assign them to one language family rather than another. These would be mixed languages.
The question is discussed in detail in, for instance, Thomason and Kaufmann (1988).

For the vast majority of the world’s languages, there is no reason to think that they are not well behaved with respect to their genealogical affiliation, i.e., even though all languages have borrowed at least something from other languages with which they have been in contact, they have a clearly identifiable core that has been inherited by direct vertical transmission from a single ancestor language. But there are some exceptions to this generalization so that the possibility of mixed languages must be acknowledged. One set of languages that belongs here consists of the creole languages of the Caribbean such as Haitian Creole and Jamaican Creole. These languages take most of their vocabulary from a European language—French in the case of Haitian, English in the case of Jamaican—but have a grammatical structure that is very different from that of these European languages and more reminiscent of the grammatical structures of the West African languages that were the native languages of the slaves brought to the Caribbean and among whom the creole languages developed. Haitian Creole, for instance, is arguably a language with French vocabulary but Fongbe grammar. (Fongbe or Fon is the major indigenous language of Benin in West Africa.)

A perhaps even more striking example is Mitchif, spoken in North Dakota and scattered locations in Canada, since in Mitchif the dividing line between the different components cuts across the divide between vocabulary and grammar. In Mitchif, most verbs are of Cree origin, Cree being a member of the Algonquian family, one of the major indigenous language families of northeastern North America. Moreover, Mitchif verbs show the complex verb morphology that is characteristic of Algonquian languages. By contrast, nouns and adjectives are overwhelmingly of French origin and show characteristics of French noun phrases that are not found in Algonquian languages such as gender agreement between prenominal adjective and noun. While there are some exceptions, such as French-origin verbs and Cree-origin nouns (especially kin terms), the basic division is robust. A Mitchif sentence is thus basically a framework created by the Cree-origin verb into which one plugs noun phrases that are composed according to the rules of (somewhat modified) French. This makes it hard to say that Mitchif is either unequivocally an Algonquian or an Indo-European language. Of course, if one wants, it is always possible to force a decision, and in the case of Mitchif this would probably be in the direction of Algonquian, given that the Cree component provides the basic clause structure into which French-like noun phrases are inserted. But this is still very different from, say, the French loans into English, since one can perfectly well construct many sentences in English that completely lack French loans, whereas it is virtually impossible in Mitchif to construct sentences containing noun phrases without resorting to at least some French-derived vocabulary and grammar.
Although mixed languages provide exceptions, or at least challenges, to the simplistic assumption that every language is necessarily assigned to one and only one family, the study of such languages has provided important insights into linguistic history and prehistory. Bakker (1997), for instance, shows how a combination of linguistic and ethnographic-historical research can uncover much of the prehistory of how and why Mitchif arose in the way in which it did.

25.2.5 Linguistic paleontology

So far, we have seen how the comparative method can be used to establish that languages are genealogically related and to reconstruct their common ancestor—at least in principle, although the only example worked through in any detail involves the reconstruction of Proto-Indo-European *p as the common ancestor of Proto-Germanic *f and the corresponding *p found in most other Indo-European languages. It should be noted that the method cannot have as a possible result that two languages are not genealogically related, although it may have as a result that no genealogical relationship can be established—lack of evidence is not evidence of lack. This is already a significant result, since it shows, for instance, that the speakers of such modern languages as English, German, Spanish, Russian, and Armenian are linguistically the descendants of a single speech community, to which we have given the name Proto-Indo-European. But there are lots of things that are left open.

First, this says nothing about the biological ancestry of these speakers, and only biological investigations, in particular population genetics, will answer this question. There is no necessary link between the transmission of genes from generation to generation, which is a biological process, and the transmission of languages from generation to generation, which is a cultural process. And while early work suggested that there might be a close link between language families and families of populations (Cavalli-Sforza et al. 1988), later work has at least drawn our attention to the possibility of discrepancies (Nasidze et al. 2004), the latter suggesting, for instance, that while the Azerbaijani language belongs to the Turkic family, its speakers belong genetically rather with other populations of the Caucasus, which in turn suggests that a language shift took place whereby an indigenous population abandoned its heritage language for the incoming Turkic language. This is, incidentally, a nice illustration of how linguistics and genetics can cooperate in unraveling the details of prehistoric events.

Second, the method as described so far says nothing about the environment in which such reconstructed languages as Proto-Indo-European were spoken, whether by environment we mean the physical or the cultural environment. If, for instance, we were to discover an archeological site that lacked overtly...
linguistic evidence, such as inscriptions, we would have no way of knowing whether this might be plausibly identified with speakers of Proto-Indo-European or not. An extension of the comparative method, so-called linguistic paleontology, does, however, provide us with a tool of just this kind. Let us suppose that by using the comparative method as described so far we have established cognates for the word expressing a particular concept across a number of languages and are therefore confident that we can project the word back to the protolanguage. Crucially, this involves not only projecting the form but also its meaning, since any word (or more accurately: linguistic sign) is a combination of form and meaning. But this means that we can reconstruct not only, to a good degree of approximation, the forms that characterized the protolanguage but also the concepts that were around when that protolanguage was spoken. This is the field of linguistic paleontology. It has been applied to a number of language families, but we will here take examples from Indo-European.

Concepts that can be reconstructed to Proto-Indo-European provide significant evidence concerning the cultural environment in which this language was spoken; wide-ranging evidence and conclusions are presented in Mallory and Adams (1997). They show a society familiar, for instance, with textiles, metallurgy, wheeled vehicles, and agriculture, indeed probably quite advanced agriculture, given that a word can be reconstructed for “plow,” although the invention of the plow postdates the introduction of agriculture by around three millennia in the ancient Near East and adjacent Europe. It is important to note that linguistic paleontology is primarily a case of positive evidence. The fact that a particular concept cannot be reconstructed to the protolanguage does mean that the concept did not exist, only that there is no evidence to show that it did—but evidence may well have been lost to such an extent that an originally present concept can no longer be reconstructed. Many Indo-European branches and languages have, for instance, created new words for “hand,” to such an extent that it is difficult or impossible to reconstruct what the word for “hand” would have been in Proto-Indo-European, although we can surely assume that the speakers of this language did have hands.

25.3 Other traditional methods and representations

25.3.1 Internal reconstruction

Although the comparative method is the most widely used and most reliable method for reconstructing linguistic prehistory, there are other methods that can sometimes be brought into play, of which one forms part of the arsenal of
traditional historical-comparative linguistics, namely internal reconstruction. In order to explain this method, it will be useful to examine a historical change that took place in the history of the German language. At the earliest relevant stage voiced and voiceless stop consonants, such as \(d\) and \(t\), contrasted word-finally, as they still do in English *made* /meid/ versus *mate* /meit/. This state of affairs is still retained in German spelling, where we have for instance a distinction between *Bund* “union” and *bunt* “multicolored.” In Old High German these two words would also have been pronounced differently, but in the modern language they are both pronounced /bunt/, although they are still distinguished when a vowel follows within the same word, as in the genitive singular *Bundes* /bundəs/ versus the neuter adjective *buntes* /buntəs/. The reason for this state of affairs is a sound change that took place in the history of the German language and devoiced word-final stop consonants, including devoicing word-final /d/ to /t/. It will be seen from the examples given that different forms of the same word can be differentially affected so that we have a voiceless /t/ in *Bund* but a voiced /d/ in its genitive singular *Bundes*. Historically, a sound change applied in a particular environment (word-finally) but not elsewhere (in particular, before a vowel). If we just look at the modern language, namely the alternation between *Bund* /bunt/ and *Bundes* /bundəs/, then we have a so-called morphophonological (morphophonemic) alternation. In other words, environmentally conditioned sound changes can give rise to morphophonological alternations.

We can now turn this process on its head in order to reconstruct earlier states of a language, by hypothesizing that where we find a morphophonological alternation in an attested language, this must result from an earlier stage where there was no such alternation and an intervening sound change that affected the sound that alternates (/d/–/t/ in our example) in one environment but not the other. Now, this is far from foolproof argumentation; logically, it is the equivalent of trying to infer \(p\) from \(\text{if } p \text{ then } q \text{ and } q\). But nonetheless, it does provide a plausible account of how a language can come, synchronically (i.e., at a given point in time), to have a morphophonological alternation.

The method can actually be generalized somewhat. Whenever we have a situation that seems synchronically to be aberrant, we can ask what plausible chain of historical changes could have led from an expected situation to the currently observed unexpected situation. While the method has its risks, it has also paid off in at least one famous case. Saussure, better known as one of the founders of structural linguistics, was aware, like many linguists before him, of the kind of vowel alternations found in some Indo-European languages, such as in the Ancient Greek verb “to leave” with present tense *leípō* “I leave,” perfect *léloípa* “I have left,” aorist *èlìpon* “I left,” with the general pattern of alternations being \(e-o-Ø\), to which the name “Ablaut” is conventionally given in
Indo-European studies. Saussure noticed that some verbs are apparently exceptional, in particular in having the vowel *a* or *o* for expected *e*. He posited that the unexpected vocalism was the result of a sound that was originally adjacent to the vowel in question, that had affected the quality of the vowel (e.g., changing *e* to *a*), and had subsequently been lost; these “lost” sounds are now conventionally referred to as laryngeals. At the time of Saussure’s publication (1879), this was considered pure conjecture, but when later the Hittite language was deciphered and shown to be Indo-European, it turned out that Hittite had the consonant *h* in many of the places where Saussure had predicted the presence of a laryngeal.

25.3.2 Wave model

The family tree model of linguistic relatedness presented in Section 25.2.1 is the basic method used in order to represent genealogical relationships among languages. Nonetheless, there is a clear sense in which it is unrealistic. It seems to say that at a particular point in time, a language splits into two languages such that those two languages no longer have anything to do with one another. Under certain specific circumstances this might be plausible, for instance, if a speech community splits into two by one part going by boat to a distant island and never returning or being followed by those left behind, but this is clearly unlikely to be the normal state of affairs. Rather, what one finds is a gradual breakup of a speech community, with differences gradually increasing and fault lines slowly developing that may, subsequently, give rise to distinct languages and, eventually, even distinct branches of a language family. But unless and until this fission finally takes place, the picture one has is not so much of a family tree as of a patchwork of differences that crosscut one another, making the construction of a single family tree unjustifiable.

An example from the history of English will illustrate the general point. Two innovations that differentiate Modern English from Middle English (the language of, for instance, Chaucer) are (1) the diphthongization of the long vowel ā to give *au* as in *house* and (2) the replacement of the inherited third person plural pronouns by forms borrowed from Old Norse and which ultimately give rise to Modern English *they*, *them*, *their*. Innovation (1) spread from the south of the English-speaking area to the north, while innovation (2) spread from the north to the south. Let us go back to the period when these were ongoing changes and take a variety of English that has undergone both changes (henceforth “Middle”) and compare it with varieties that have just undergone change (1) (henceforth “Southern”) or just undergone change (2) (henceforth “Northern”). Since both Middle and Southern have undergone change (1), they should be
grouped together under a single node of the family tree, and this node should not include Northern, which has not undergone change (1). But conversely, since both Middle and Northern have undergone change (2), they should be grouped together under a single node of the family tree to the exclusion of Southern, given that Southern does not share this innovation. But this leads to a contradiction, since there is no way to draw a well-formed tree that satisfies both requirements. Rather, the results of the innovations can be diagramed as waves emanating from a particular starting point (the south for (1), the north for (2)) and overlapping in the middle.

Wave models are particularly useful in diagramming relations among closely related language varieties that are still in contact, and maps of dialects of a single language are often a good illustration of the use to which the wave model can be put. (It should, however, be noted that dialect maps often just show the geographical distribution of different variants without indicating which is the innovation, thus abstracting away from a representation of historical change.) The same can be true of closely related varieties that are considered distinct languages, often for reasons as much political or social as linguistic, if not more so. More generally, wave diagrams provide a more accurate model of reality “on the ground” as a linguistic area is in the process of undergoing internal differentiation. Subsequent changes, in particular, the loss of intermediate varieties as more prestigious varieties spread, will typically later override this picture to give clearer boundaries and a situation that can ultimately be diagramed by means of family trees; a more detailed discussion of how this can happen, with particular reference to the differentiation of the major branches of the Indo-European family, is given in Garrett (1999). But even after this differentiation has taken place, there can still be residual features that override the neat representation implied by the family tree. Within Indo-European, for instance, Slavic languages share some features with Germanic, other features with Indo-Iranian, although there is no overall subgrouping defined by shared innovations that includes all of Slavic, Germanic, and Indo-Iranian. More generally, the family tree model, with its implication of strict parthenogenesis, is a historical simplification that is extremely useful but should nonetheless not be elevated to the level of absolute truth.

25.4 Shortcuts and other new methods

25.4.1 Mass comparison

As will be appreciated from the discussion so far, traditional methods in historical linguistics require detailed investigation of the languages involved and the
painstaking construction and testing of hypotheses concerning the historical relations among them. Historical linguists will therefore often devote considerable effort to, for instance, the etymology of a particular word, arguing whether it is inherited from a particular ancestor language or a loan from outside, as evidence in arguing for genealogical relatedness among languages or for or against a particular subgrouping. It is therefore inevitably tempting to ask whether there are not shortcuts that could produce the general results that we are seeking without requiring us to go into all the detail of, for instance, trying to reconstruct the history of the pronoun *she* in English. One such method, championed by Greenberg (see, for instance, chapter 1 of Greenberg 1987) is mass comparison, which works by looking for similarities across a large number of languages. The method requires the compilation of comparative word lists, but does not require the establishment of regular sound correspondences, and thus does away with much of the apparatus required by the mainstream comparative method. Unfortunately, this also means that one loses the checks that this comparative method imposes, against, for instance, the inclusion of accidental look-alikes of the English *much*–Spanish *mucho* type, or the exclusion of regular correspondences that happen not to involve similar forms, as with Armenian *erku*–Russian *dva* “two.” While the method has been widely touted as a way of establishing more wide-ranging genealogical groupings than is possible using the comparative method, the results have not met with widespread approval, and it is perhaps unfortunate that they have too often been taken on faith by specialists in other disciplines.

But the attraction of mass comparison does point to a weakness of the comparative method, one that might or might not be able to be overcome by other methods. The accumulation of gaps and statistical inaccuracies as one reconstructs back in time means that eventually a point is reached where one can no longer be sure that such similarities as one detects between languages are the result of common ancestry rather than attributable to other factors, in particular chance and borrowing. It is hard to pin a date on this point—as will be noted in \textsection 25.4.2, there may well be no aspect of language structure that changes with the regularity of radioactive decay or the molecular clock—but it is empirically true that the languages families that have gained ready acceptance among linguists do not go back beyond about 10,000 years BP. Various factors can affect the exact date: for instance, one can usually go back further in time with a family of languages like Indo-European, which includes a large number of languages (about 200 on a fairly conservative count) with an internal division into a number of well-defined branches, than with a family that includes few languages or has little internal diversification. Nonetheless, it remains the case that most linguists are skeptical of claims going back further than around
10,000 years BP (and of a number of claims that do not go back that far), which is one of the reasons why several linguists are searching for new methods that might enable us to go beyond this limit.

### 25.4.2 Glottochronology

Although the comparative method leads to the construction of hierarchical tree diagrams that mirror, albeit with some simplification (as discussed in Section 25.3.2), the historical development of the descent of languages from common ancestors, the method does not in and of itself include any absolute chronology, so that the claim that English, German, and Swedish all descend from Proto-Germanic, and that Proto-Germanic and Proto-Slavic both descend from Proto-Indo-European, gives us no indication when Proto-Germanic, Proto-Slavic, or Proto-Indo-European was spoken. While linguists have not necessarily been reluctant to assign dates to such ancestor languages, this assignment has traditionally been intuitive, based on a “feeling” about how quickly languages change to produce particular degrees of inter-linguistic difference. For instance, Indo-Europeanists usually assign a date of about 6,000 years BP to Proto-Indo-European, but it is difficult or impossible to produce hard evidence for this date, although dates widely divergent from this can be considered implausible because, for instance, they would require language change during the early history of the family to have been much slower than the attested rate of change of any known language in the historical period—one reason for linguists’ skepticism toward dates for Proto-Indo-European of around 9,000 years BP.

In some disciplines there are methods of establishing quite reliable chronologies by making use of changes that are known to occur at fixed rates, such as the decay of radioactive materials in radiocarbon dating and other comparable techniques in archeology, or the molecular clock in genetics. Linguists have therefore sought after a comparable measuring stick in linguistic material. The method that has gained most widespread, though still quite limited, support is glottochronology. (A summary of both traditional and more innovative versions of glottochronology is given in Starostin (2000), which also includes criticism of traditional glottochronology and quite far-reaching proposals for modifications to the method to achieve greater reliability—modifications that still require testing.) The basic assumption of glottochronology is that for a certain portion of the vocabulary, the so-called basic vocabulary, there is a fixed rate of replacement, such that, for instance, after a thousand years 14% of this basic vocabulary will have been replaced. The choice of vocabulary might initially seem strange, since borrowing words from one language into another is perhaps the most
obvious sign of language contact. However, borrowing applies primarily to the nonbasic, culturally laden vocabulary, so that if one examines, for instance, Latin and French loans into English, there are only five in the 100-word “Swadesh list” (*animal, grease, mountain, person, round*), although overall English has borrowed substantially from these sources. The basic vocabulary, incidentally, is designed to include only terms that are part of basic universal human experience, though occasional infelicities creep into lists, such as “ice,” which does not occur naturally in most tropical regions, or “horn”—there are no indigenous horned animals in New Guinea.

One problem with the general method is that, in contrast to the situation with radioactive decay or genetic mutation, there is no principled reason to assume that the rate of replacement of basic vocabulary should be constant, in particular independent of social factors, which are known from experience to be the main determinants of the linguistic effects of language contact. Indeed, there may be particular circumstances in which social factors determine unusual rates of lexical replacement, even in basic vocabulary. Comrie (2000) discusses the case of the Haruai language of Highland New Guinea. In Haruai society, most personal names are ordinary words, and speakers are under a taboo not to utter the names of certain kin (in-laws, and father’s sister’s or mother’s brother’s children) or the ordinary words identical to these names, but rather to replace them with a synonymous lexical expression. Given the small size of the community, with about 1,000 speakers, this leads to a situation where words can go out of use quite rapidly because of broad kin networks relative to the small size of the speech community, typically being replaced by loans from adjacent languages. This has given rise to a situation where Haruai shares about the same percentage of basic vocabulary with Hagahai, to which it is genealogically related, as with Kobon, to which it is not demonstrably related. Starostin (2000) discusses a number of incorrect estimates of time depth that have been reached using glottochronology, at least in its traditional form.

There are various claims that have been made for glottochronology, with corresponding differences in how exactly the method is applied and interpreted. One implementation would search for look-alikes across languages, without considering in detail whether the words can actually be shown to be cognate (to descend from a common ancestor), calculate the percentage of such items across the whole list, and from this calculate the closeness of the genealogical relationship between the languages, and the date at which their common ancestor was spoken, always bearing in mind that very low percentages may be indistinguishable from chance similarity. This implementation suffers from the disadvantages of mass comparison discussed in Section 25.4.1. A more strict interpretation would be to calculate only the percentage of cognates—which
presupposes that one already knows that the languages are genealogically related—and then use this to calculate the time depths at which particular languages in the family separated off from other languages and therefore to draw a family tree with chronological labeling of the individual nodes. But all of this rests on the debatable notion of a fixed rate of change.

25.4.3 Typological comparison

The method of typological comparison has been championed by Nichols, for instance in Nichols (1992). A typological similarity across languages means a similarity not in form but rather in the overall structure of some part of the language. For instance, some languages like French, German, Russian, and Swahili have gender systems, whereby nouns are divided into two or more classes (genders) and other items, such as adjectives, have to change their form (agree) according to the gender of the noun with which they are associated. In French, for instance, the noun *fils* “boy” is masculine, and therefore requires a masculine definite article and adjective in *le bon fils* “the good son,” whereas *fille* “daughter” is feminine and requires a feminine definite article and adjective in *la bonne fille* “the good daughter.” By contrast, other languages, such as Hungarian and Persian, lack a gender system. (English, incidentally, does have a gender system, but with rather minimal scope: the third person singular pronouns *he, she, it* are selected in agreement with the noun to which they refer.) As can be seen in this example, typological classifications do not necessarily correspond to genealogical families: French, German, Russian, and Persian are all Indo-European, while Swahili and Hungarian are not.

The challenge posed by typological comparison as a tool in comparative-historical linguistics is to find typological features that are highly resistant to change, and are thus likely to remain constant across the languages of a particular family. If such features can be found that are even more persistent than the chronological limits on the comparative method, then we would be enabled to go further back into prehistory than is possible with that method.

Overall, the general result of typological comparison is that most typological features are very sensitive to language contact and thus cannot serve as indicators of genealogical relatedness. Indeed, one of the main results of the massive areal typological project reported in Haspelmath et al. (2005) is that languages are frequently more similar typologically to their neighbors than to their relatives. While some of the particular features proposed by Nichols have been criticized on the basis that they are subject to change by language contact, this does not exclude the possibility that there might be some typological
features that can be used as indicators of deep genealogical relations among languages.

One that has been proposed by Nichols is precisely the existence of gender systems, although the criterion is more complex than the mere presence versus absence of a gender system in a language. The absence of a gender system is not a plausible criterion for deep genealogical relatedness, since we know that languages can lose their gender systems, as has happened in the case of a number of Indo-European languages, such as Persian and Bengali (both belonging to the Indo-Iranian branch of the family, though almost at opposite western and eastern geographical extremes, respectively, of that branch), and Armenian. Moreover, languages can develop gender systems under contact with other languages that have gender systems, as has happened in a number of indigenous languages of Australia. But what may be more difficult is for a language spontaneously to develop a gender system. If this is so, then it would suggest that if a language that has a gender system is spoken in an area where other languages lack gender systems, then the genealogical relatives of that language should be sought with other languages having (or at least: having had) gender systems. A case in point would be Ket, the sole surviving language of the small Yeniseic family of western Siberia, which has a gender system, although all surrounding languages do not (these languages belong to the Uralic, Turkic, and Tungusic families—Russian does have a gender system but is a very recent intruder into this geographical space). The more extensive testing of Nichols’ specific claims about particular typological features remains a task for future research.

25.5 Conclusions

Linguistics has a distinguished history of establishing genealogical relationships among languages and thus providing evidence for the common linguistic ancestry of apparently diverse speech communities, as well as of reconstructing aspects of the culture and physical environment in which the speakers of the common ancestor language must have lived. The traditional methods require detailed and painstaking analysis of linguistic material and give at best questionable results when pushed back beyond 10,000 years BP (and in many cases well short of that range). New methods have been proposed that would enable linguists to go beyond the 10,000-year range, as well as to establish more reliable chronologies, but unfortunately these methods all suffer from questionable assumptions or proven inadequacies that have left most linguists skeptical of their validity. Nonetheless, interdisciplinary investigations involving linguistics, archeology,
and genetics have succeeded in throwing light on what until recently seemed to be intractable problems.

References


Saussure F (1879) Mémoire sur le système primitif des voyelles dans les langues indo-européennes. Teubner, Leipzig


26 General Principles of Evolutionary Morphology

Gabriele A. Macho

Abstract

Anthropologists analyzing morphology for phylogenetic, functional, or behavioral purposes are confronted by a plethora of obstacles. Morphology is not free to vary but is subject to a number of constraints, which may be historical, developmental, and/or functional. This, together with the fact that the fossil record is scant, confounds meaningful interpretation of phylogenetic pathways and the reconstruction of function and behavior from fossilized remains. To overcome these difficulties, paleoanthropology is becoming increasingly inter- and multidisciplinary, whereby researchers draw on, and incorporate, approaches and findings obtained in other, sometimes very diverse disciplines. The purpose of this contribution is to review the constraints acting on morphology, the limitations faced when interpreting form/function and behavior from morphology, and the different approaches currently explored in paleoanthropology.

26.1 Introduction

With rare exceptions (e.g., endocasts, footprints, permafrost remains, peat bodies), only the hard tissues of the body become fossilized. Phylogenetic, functional, and behavioral interpretations of extinct taxa are thus almost exclusively based on analyses of bones and teeth. Such fossil remains are often incomplete and distorted, while taphonomic biases prevent a good representation of all parts of the body and/or functional units for detailed analyses (Grupe Volume 1 Chapter 7). It is therefore not surprising that the search for more fossil material is a priority in paleoanthropology (Brunet et al. 2002). These shortcomings aside, there exist other, more fundamental problems in analyzing and interpreting morphology. Resolving these issues is the aim of diverse research areas, whose integration has only just begun.

Morphology is highly constrained and determined by a cascade of interactions between genetic, epigenetic, and environmental factors modulated through past and present selective pressures (Atchley and Hall 1991). Consequently,
Morphology and function must not only be interpreted with respect to present-day function and the fitness it may confer to the species but should also be viewed against the backdrop of past form, function, and selective pressures. Phenotypic correlations due to pleiotropy, epigenetics or correlated selection, genetic drift, and co-option of characters for functions they were not initially selected for further compound a sound appraisal of morphology within a phylogenetic and functional framework. This, together with the fact that the developmental pathways for various structures are only poorly (if at all) understood, raises the question of whether form follows function or whether the reverse is just as likely. This is problematic as speciation is commonly associated with the acquisition of novel morphological characters and/or the exploitation of new ecological niches (Müller and Wagner 1991; Shubin 2002; but see Groves Volume 3 Chapter 18).

To overcome problems associated with the interpretation of morphological structures and variation in both extant and extinct populations, evolutionary studies have become increasingly multi- and interdisciplinary, although there are two main strands of investigations. First, evolutionary developmental biology, despite its long (and sometimes misguided) history (Haeckel 1866), is now a well-established discipline (Hall 1992; Wagner and Gauthier 1999; Love and Raff 2003). Such studies combine an experimental, developmental research protocol with comparative analyses of the fossil record to determine the developmental and phylogenetic pathways of morphological structures and the constraints acting on their formation. Despite their high power of resolution, such studies require expert knowledge, are time consuming, and are financially costly. Hence, they are restricted to a few laboratories and are not directly amenable to traditional paleoanthropological enquiry. Second, new functional and biomechanical tools are being incorporated in paleoanthropological research protocols (Thomason 1997) and promise to make a contribution to our understanding of the functional adaptations of extinct species.

A comprehensive review of both strands of enquiry is beyond the scope of this contribution. Instead, this chapter aims to highlight general problems associated with interpreting the morphology observed in fossil remains and the ways in which paleoanthropologists are currently investigating constraints, function, and behavior in the fossil record using a predominantly comparative approach.

### 26.2 Comparative morphology and constraints

Morphology is not free to vary but is subject to a number of developmental, physical, and historical constraints (Maynard Smith et al. 1985), also categorized
as generative or selective constraints (Richardson and Chipman 2003) (Figure 26.1). As the field of developmental biology progresses, new classes of genetic linkage constraints are being identified (Calabretta et al. 2003), which complicates the elucidation of evolutionary pathways even further. Yet, anthropological research is restricted to determining constraints through (1) quantitative methods, (2) by indirectly taking recourse to developmental pathways, and (3) through research designs that incorporate some (direct or natural) functional experiments.

### 26.2.1 Comparative morphology and evolutionary constraints

Phylogenetic systematics or cladistics is the main tool for reconstructing the phylogenetic relationships of taxa (Hennig 1966) and should precede paleobiological enquiry and functional analyses. While the strengths and weaknesses of this method, as well as its assumptions, are reviewed elsewhere in this handbook (Chapters 4–6) some limitations pertaining to the analysis and interpretation of morphology need to be briefly reiterated. Relevant for the discussion of constraints are issues regarding homology and the independence of characters. Analyses of homologous traits are at the heart of cladistic methodology, but experimental studies have provided unequivocal evidence for the hierarchical nature of homology, whereby homologies at one level of organization need not lead to homologies at another (Hall 1994; Raff 1996; Wagner and Gauthier 1999). Perhaps even more problematic for broader morphological enquiry is the assumption that anatomical traits are independent. This clearly is not the case and many characters are correlated as a result of pleiotropic effects (Riska and Atchley
1985; Riska 1989), correlated selection (Leroi et al. 1994), modular networks (Kauffman 1993; Salazar-Cuidad et al. 2000; Salazar-Cuidad and Jernvall 2002), and genetic linkage (Kangas et al. 2004). Such intercorrelations not only obscure the phylogenetic history but also hamper inferences made about function and behavior (see later). Determination of the covariance of various traits (Atchley and Hall 1991) is therefore of concern to paleoanthropologists.

On the basis of methodologies developed in population genetics, various quantitative tests are available to appraise the covariation and correlation of traits with the aim of inferring developmental and phylogenetic constraints (Cheverud et al. 1989). Such approaches are becoming increasingly popular in paleoanthropology and have recently been extended to the study of skull variation (Ackerman and Cheverud, 2000; Lieberman et al. 2004) and teeth (Hlusko et al. 2004). Such research designs not only shed light on the correlation of characters but also have the potential to estimate the rate at which characters can change during evolution. For example, Hlusko et al. (2004) have estimated that enamel thickness in baboons could theoretically double in only 50,000 generations. If correct, this character could be prone to homoplasy, and enamel thickness as a defining trait in hominins should be interpreted with caution.

Where more behavioral aspects of morphology are the focus of study, independent contrast, pioneered by Felsenstein (1985), is often the method of choice. Using a relatively simple algorithm, the independence (or otherwise) of two traits can be assessed. While originally used in zoological and behavioral research, recent paleoanthropological studies have begun to use this method also (MacLeod et al. 2003). Independent contrast is a powerful tool in evolutionary studies provided that the phylogenetic history of species is known: morphological traits under investigation must be mapped onto phylogenies. However, given that the primate phylogenies as well as their divergence times are subject of research themselves, the results obtained from such analyses should only be regarded as working hypotheses rather than providing conclusive information about the coevolution of morphological traits.

### 26.2.2 Comparative morphology and development constraints

#### 26.2.2.1 Teeth

Teeth are among the most abundant remains within the fossil record and contain a wealth of information for the taxonomists (Bown and Rose 1987; Haile-Selassie et al. 2004) and the functional morphologist (Hiiemae and Kay 1972; Kay 1977).
Owing to their pivotal role in the breakdown of food (Strait 1997), overall tooth size, shape, and enamel thickness are considered good indicators of the dietary niches exploited by extinct species (Janis and Fortelius 1988) and, when analyzed across lineages and clades, may even provide information about global climatic fluctuations (Fortelius et al. 2002). The importance of tooth morphology for functional inferences is thus undisputed, but the high level of homoplasy resulting from this fact compromises the usefulness of teeth for taxonomic purposes (Teaford et al. 2000). For example, the hypocone probably evolved independently more than 20 times (Hunter and Jernvall 1995), while thick enamel among primates could be the result of convergent evolution also (Janis and Fortelius 1988; Hlusko et al. 2004). Nonetheless, while overall tooth size and shape exhibits a considerable amount of homoplasy, the way in which these morphologies are achieved appears developmentally constrained.

Teeth are unique biological structures that retain a permanent record of aspects of their ontogeny, i.e., timing and rates of formation (Macho and Wood, 1995), as well as cell movement during odontogenesis (von Koenigswald and Sander 1997). Enamel and dentine matrix is secreted in a circadian manner during odontogenesis, whereby the daily increments are manifest in the mineralized tissue as cross-striations (enamel) and von Ebner lines (dentine). Analyses of these incremental lines in histological sections of teeth (or their surface manifestations) have allowed an appraisal of the timing of tooth development across extant and extinct primates (Bromage and Dean 1985). Apparently, dental crown formation times in primates are highly correlated with brain size and, consequently, other aspects of life history (Macho 2001; see also Hemmer Volume 1 Chapter 18), probably as a result of pleiotropic effects (Riska and Atchley 1985; Macho 2001; see also Hemmer Volume 1 Chapter 18). This mirrors restrictions observed in the timing and pattern of tooth eruption in primates (Smith 1989), as well as across mammals (Smith 2000). Thus overall analyses of dental development, both through histological sections of isolated teeth and eruption patterns, allow to address broader questions of brain and life history evolution, as well as developmental constraints. In addition, detailed analyses of the rates of enamel secretion (Beynon et al. 1991) could make possible an assessment of the heterochronic changes that occurred during dental evolution (McNamara 1995), although a satisfactory method for ascertaining the amount of enamel secreted by an ameloblast per day is still wanting.

Enamel prisms are formed by ameloblasts and trace the movement of ameloblasts during odontogenesis. While the mechanisms underlying the evolution of prismatic enamel, a characteristic trait of eutheria (von Koenigswald and Clemens 1992), are still the subject of investigation (von Koenigswald and Sander 1997), once evolved, prismatic enamel seems to have provided great potential for
further evolutionary modifications in response to selection pressures. As a case in point, the crossing-over of bundles of prisms, i.e., prism decussation, most probably evolved in large-bodied, thick-enamelled species as a crack-stopping mechanism (von Koenigswald et al. 1987). Until recently however, visualization and quantification of the complex three-dimensional arrangement of prisms has been wanting (Jiang et al. 2003; Macho et al. 2003a). Preliminary studies using this new technology indicate that, despite similarities in prism undulations, there are fundamental differences in the pattern of cell movement and the amount of prism deviation among even closely related primates (Jiang et al. 2003; Macho et al. 2003a). If confirmed, this could be exploited further for taxonomic and functional purposes (Rensberger 2000; Macho et al. 2005; see later).

26.2.2.2 Bone

Unlike enamel, bone is a plastic material, which continues to remodel throughout life in response to loads placed on it so that size and structure provide information about the functional adaptations of bone (Wolff 1892). For example, joints have evolved to confer both optimal orientation with regard to the direction of load and mobility, although the relative contribution of each of these aspects, as well as the magnitude of loads, remains uncertain. Cortical bone provides bending strength to the shafts of long bones. Overall scaling relationships between bone size and body mass can thus be explained (Jungers 1985). Furthermore, weight-bearing bones (e.g., vertebrae, calcaneus) or weight-bearing aspects of long bones (i.e., proximal ends) gain their structural stability through a meshwork of interconnected trabeculae. Such an arrangement ensures maximum strength while at the same time minimizing weight; unsurprisingly 70% of all the bone is thus trabecular bone (Huiskes 2000). In order to optimize the functional adaptations of bone, trabeculae tend to align along the principal stress trajectories (Biewener et al. 1996; Gefen and Seliktar 2004). While these general relationships appear valid, studies have questioned whether bone is indeed as plastic, i.e., responsive to loading conditions, as previously thought, especially after adolescence (for recent review see Pearson and Lieberman 2004).

In an innovative comparative study, Ruff et al. (1994) investigated the response of bone remodeling during different stages of ontogeny. They found that increased mechanical loading prior to (or during) adolescence leads to an increase of cortical bone thickness due to periosteal expansion. Conversely, increased loading during the later stages of development results in endosteal contraction with external diameters remaining relatively unaltered (Ruff et al.
Furthermore, cortical bone responds systemically to increased activity levels during ontogeny, even in the absence of direct loading (Lieberman 1996). Increase in bone length, on the other hand, is largely stimulated through the effects of growth hormone at the growth plates and via muscle action (Vogl et al. 1993). Thus, in summary, it would seem that physical activity during development, i.e., during adolescence in particular, will have the greatest effect on bone remodeling, whereas activity levels during adulthood will result in only minor changes. Reconstruction of population-specific activity levels from adult bone geometry alone could therefore be misleading (Macho 1990, 1991; Ruff et al. 1993) and results need to be interpreted with caution. Similarly, trabecular alignment within bones is apparently developed relatively early during development (Biewener et al. 1996; Tanck et al. 2001) and is the result of differential growth in areas where strengthening is needed and resorption, especially later in life, where it is not (Biewener et al. 1996; Tanck et al. 2001). However, and unlike cortical bone which will only remodel in adulthood, trabeculae retain some potential to model throughout life (Frost 1990; Kobayashi et al. 2003). This is an exciting aspect of trabecular bone, and anthropological studies have already begun to exploit the functional information contained within cancellous bone in extant primates and extinct hominins (Macchiarelli et al. 1999; Fajardo and Müller 2001; Fajardo et al. 2002). While trabecular orientation may prove indicative of the loading directions experienced during life, much elucidation of phylogenetic constraints may come from analyses of carpals and tarsals.

Carpals and tarsals lack epiphyseal growth plate and, with some rare exceptions, also lack muscle attachments (Lewis 1989). Prior to ossification, these bones exist as cartilaginous models, resembling the future bone shapes. While the articular cartilages have a germinal layer, articular cartilage chondrocytes are much less responsive to growth factors than those of epiphyseal plates (Moss-Salentijn 1992). Owing to the absence of both growth plates and muscle attachments, carpals and tarsals therefore seem less influenced by changes in growth hormone (unlike long bones and cortical thickness of the skull). Rather their size and shape is apparently more strongly determined by intrinsic mechanisms than long bones (Bryant and Simpson 1984; Dainton and Macho 1999a, b). If correct, the mosaic nature of morphological changes seen in the hominin fossil record can be explained. For example, Australopithecus anamensis exhibits unequivocal bipedal traits in its lower limb morphology, but the carpal morphology resembles that of the chimpanzees (Leakey et al. 1998; Ward et al. 2001). Although it is recognized that this may indicate retained adaptations to arboreal life (Leakey et al. 1998; Ward et al. 2001), based on the results of the studies outlined earlier, it is not unreasonable to suggest that changes in wrist bone morphologies may lag behind those of long bones due to phylogenetic constraints.
26.3 Determination of functional adaptations

Adaptations are traits which enhance the fitness of the species and have arisen over evolutionary time as a result of natural selection for present biological roles (Rose and Lauder, 1996). Inferring the functional adaptations is not trivial, however, even when the traits under investigation do not constitute evolutionary novelties. Comparative analyses based on analogy with extant taxa assume that structure and function are intricately linked, but this is seldom the case (Lauder 1997; see later). Furthermore, most biological structures have evolved to fulfill more than one function. The role of the primate hand in locomotion and manipulation (Fleagle 1999) and the human orofacial skeleton in mastication and communication (Shiller et al. 2002) are only two of many examples.

From a phylogenetic perspective, traits which are advantageous and enhance the fitness of the species may in fact have evolved for a different purpose or they could initially have been the result of random genetic drift, pleiotropy, or correlation with other structures (Gould and Vrba 1982). In order to affirm that evolution has selected for a certain trait, it is imperative that extant analogs exist and that the function is similar in all living species. Inferences about functional adaptations thus formulated should then be supported by biomechanical analyses (Thomason 1997). Additionally, the first appearance of the feature must be found to coincide with (and be supported by) ecological and environmental evidence, for which the feature and its functional adaptation had apparently been selected (Lauder 1982; Anthony and Kay 1993; van Valkenburgh 1994). Given the latter premise, it is not surprising that hominin morphology is increasingly interpreted within the ecological and environmental settings of these species (Vrba et al. 1995; Bromage and Schrenk 1999; Leakey et al. 2001; Bonnefille et al. 2004).

It is evident that determination of whether or not a morphological feature constitutes a functional adaptation to certain environmental conditions is not straightforward, as the fossil record is incomplete, experimental data on extant species are scarce, and there is much to learn about the environment in which early hominins evolved (Agusti Volume 2 Chapter 6; Vrba Volume 3 Chapter 4). Conversely, ascertainment of the kinematic possibilities of joint surfaces (Tocheri et al. 2003) on the one hand, and the properties of morphological structures of bone and teeth (Macho et al. 2005), on the other, may provide the necessary prerequisites for more sophisticated functional analyses and more plausible behavioral interpretations.

Over the last decade, engineering tools such as finite element stress analyses (FESA) have become incorporated in the paleoanthropological tool-kit (Ross 2005). FESA is a numerical modeling technique that examines the deformation
of a model composed of a meshwork of elements with given material properties. This technique is particularly useful for an assessment of the internal mechanical behavior of structures that cannot be tested directly (e.g., through strain gauges). The behavior of bone (Korioth et al. 1992) and teeth (Spears and Macho 1998; Macho and Spears 1999) can thus be tested noninvasively; this makes this technique particularly suited for the analyses of valuable fossil remains. Recent advances in technology have further facilitated application of this method to the study of stress flow within trabeculae (van Rietbergen et al. 1995; Gefen and Seliktar 2004). This promises to open exciting new prospects for the study of fossil hominins and the functional adaptations of their morphological remains.

As regards dental remains, FESA has already proved useful in determining the mechanical adaptations of intact teeth and has confirmed the importance of understanding the interplay between tooth loads and tooth behavior when assessing their functional morphologies (Spears and Macho 1998; Macho and Spears 1999). Specifically, the systematic results from molars of modern humans and, to a lesser extent, Pan and Pongo imply a modified mortar-and-pestle system of mastication in hominoids, whereby buccal cusps of mandibular molars may act as pestles and the occlusal basin of maxillary molars apparently provide the basin against which the food is crushed (Spears and Macho 1998; Macho and Spears 1999). Despite the limitations of the models, their usefulness for paleoanthropology is clearly evident, although the application of FESA to fossil remains provides additional challenges as the findings cannot be validated against behavioral data (Macho et al. 2005; see also in a later section).

### 26.4 Determination of behavior

This is probably the most difficult and contentious aspect of anthropological enquiry. Generally inadequate correspondence between morphology, performance, and behavior has already been highlighted (Lauder 1997). The tools available to make behavioral inferences are varied but largely rest on comparative approaches aiming to deduce the biological role of morphology (bone and tooth size and shape, muscle markings, etc.) from modern analogs. Yet, fossil species are unique and may not have modern analogs.

Over the last few years, FESA models have become more elaborate (Ross 2005) with the ultimate goal of determining the masticatory performance in extinct species. As the limitations of FESA are numerous, other techniques are more likely to make a contribution to our understanding of the functioning of complex systems. Specifically, the boom in experimental research on different aspects of biomechanics in primates (Demes et al. 1999; Hylander et al. 2000;
Ravosa et al. 2000; Schmitt and Hanna 2004; Franz et al. 2005), the application of biomechanical principles to the primate skeleton (Preuschoft 1973; Weishampel 1993), new analytical techniques (Osborn 1996; Langenbach and Hannam 1999; Koolstra and van Eijden 2001; Langenbach et al. 2002; Nagano et al. 2005), and a better, although still very incomplete, understanding of the complex networks in motor control (Ostry and Feldman 2003) have led to the application of sophisticated dynamic modeling approaches to the locomotor system (Nagano et al. 2005) and the masticatory apparatus (Langenbach et al. 2002). In extant species, such modeling approaches help to make predictions about structures that are inaccessible to direct experimentation. For example, EMG data on the pterygoid muscle cannot be obtained in vivo, except in exceptional circumstances and with great difficulty (Huang et al. 2005). Similarly, the loading conditions within the temporomandibular joint during normal mastication are inaccessible to direct measurements. As such, dynamic models are designed to formulate hypotheses about the biomechanical behavior of functional units. Although this is also true for the analyses of fossil remains, the assumptions made and limitations necessary to create the dynamic models weigh even more heavily.

First, hominin fossils are fragmentary and associated partial skeletons are extremely rare (Johanson et al. 1978, 1982, 1987; Walker and Leakey 1993; Toussaint et al. 2003). Reconstruction of the bony elements must therefore rely on the experience of the morphologist working with comparative material and/or composite constructs of functional units (Alba et al. 2003). Inherent in this latter approach is the possibility that morphological differences between individuals, as well as temporal changes in morphology, will become obscured. Second, the soft tissue and its properties can never be known, yet they must be inputted into sophisticated dynamic models. In light of recent research, which demonstrated the species-specific nature of muscle architecture (Taylor and Vinyard 2004), assumptions often made with regard to muscle properties in extinct species may be questionable. Ontogenetic changes in muscle fiber properties (Shida et al. 2005) are another confounding factor when modeling juvenile specimens and when reconstructing the life histories of extinct species. Third, the musculo-neurological network provides a feedback mechanism for movement and hence fine-tunes the functioning of complex systems (Ostry and Feldman, 2003; Malfait and Ostry 2004), but it cannot be known (and can never be tested) in extinct species. Although comparative experimental studies may ultimately determine similarities between closely related species in neural control, given the difficulties in setting up and executing such experiments, these data are unlikely to become available in the near future. What is more, even if these shortcomings could be overcome, there are other problems with which paleoanthropologists aiming to reconstruct the behavior of extinct species are faced.
The fact that a species was capable of certain functions and behaviors does not necessarily imply that it habitually did so. Both modeling and biomechanical analyses can only determine the boundary conditions of morphological structures and infer their capabilities. A distinction must be made between what a species could do and what it did do. As a case in point, a FESA study on A. anamensis dental microstructure suggests that this hominin probably dissipated higher loads during mastication than any of the extant hominoids studied (Macho et al. 2005). Whether A. anamensis habitually ate hard and tough foods or whether it was adapted to cope with the vicissitudes of annual food shortage associated with a seasonal habitat (Macho et al. 2003b) remains uncertain. It is conceivable that deterioration of the quality of food for a short period during the annual cycle could have equally selected for morphological changes as much as would a habitual shift in diet and the exploitation of a new dietary niche. Interpretation is further hampered by the fact that the effects of partitioning of the environment among sympatric species and its limiting effects on a species’ behavioral repertoire cannot be appraised from the fossil record. Among hominoids, for example, sympatric Pan and Gorilla feed on the same fruits during the fruiting season, but food selection changes more dramatically among gorillas than chimpanzees during the dry season (Kuroda et al. 1996; Yamagiwa et al. 1996). This may similarly have been the case for sympatric hominins: the overlap in morphological features and life histories between Homo and Paranthropus in East Africa (Wood and Strait 2004) could thus be explained.

Finally, there are different ways of achieving the same end-result. Marzke has devoted her life’s work to understanding the morphological correlates between hand morphology and tool use, with some encouraging results (Marzke 1971, 1997; Marzke and Marzke 1987, 2000; Marzke et al. 1992). In a recent collaborative study (Tocheri et al. 2003), she used morphometric techniques to analyze the orientation of joint surfaces of wrist bones (i.e., trapezium) in A. afarensis and Homo habilis, as well as in extant hominoids. Unexpectedly, A. afarensis appears to be similar to modern humans in joint orientation and morphology, whereas H. habilis falls within the Gorilla range. The kinematic differences among wrist bones would have affected the functional capabilities of the hands of these two species. This finding is apparently in conflict with archeological evidence of stone tool manufacture in H. habilis and led the authors to entertain the possibility that H. habilis may have manipulated the stones in ways not found in modern analogs. This need not be reflected in the end-product, i.e., stone tools. Such propositions are not only reasonable but likely and are strengthened by recent behavioral studies, which have demonstrated tool manufacture by various animals in the wild (Boesch and Boesch 1990; Emery and Clayton 2004), irrespective of morphological differences. Phrased differently, the inference of behavior from
morphology, such as tool manufacture from hand and wrist bones, seems more remote than ever.

26.5 Conclusions

Fossil remains of hard tissue, such as teeth and bones, provide a wealth of information for the evolutionary biologist aiming to reconstruct the phylogenetic history and functional adaptations of extinct species. Unfortunately, disentangling these various aspects is not trivial and requires that paleoanthropologists employ a number of analytical tools, draw on information obtained in other disciplines (i.e., developmental and experimental studies), and collaborate with researchers in other areas (e.g., engineers). Regardless, problems will remain as many of the techniques have inherent limitations and are built on assumptions that cannot be tested using the fossil record. Nevertheless, given the increasingly multi- and interdisciplinary nature of paleoanthropological research, new and exciting (as well as partially testable) hypotheses are being generated at a faster rate than ever before. This makes paleoanthropology a dynamic and challenging area of research.

References

Ackermann RR, Cheverud JM (2000) Pheno-


General principles of evolutionary morphology

26

781


Huiskes R (2000) If bone is the answer, then what is the question? J Anat 197: 145–156
Macho GA (1990) Is sexual dimorphism in the femur a ‘population specific phenomenon’? Z Morph Anthropol 78: 229–242
Preuschoft H (1973) Functional anatomy of the upper extremity. Chimpanzee 6: 34–120
Wagner GP, Gauthier JA (1999) 1,2,3 = 2,3,4: A solution to the problem of the homology of the digits in the avian hand. Proc Natl Acad Sci 96: 5111–5116
Abstract

Reconstruction plays a crucial part in working with fossils. A new aspect of reconstruction has been introduced by recent advances in 3D technology: the digital reconstruction of real objects. With the assistance of various computer-based techniques, such as surface scanning or computed tomography (CT), digital object data can be acquired and processed. Computer-based methods have in common that they are nondestructive as the handling of the fossil takes place within the computerized environment only. This aspect alone is of great advantage in comparison to conventional reconstruction and measuring techniques. Additionally, when reconstructing a fossil digitally, all steps are comprehensible and can be reversed.

With the assistance of a new methodological inventory, much additional information can now be derived from fossil specimens. Medical imaging techniques enable the visualization of internal morphological structures, formerly inaccessible for morphometric and biomechanical analysis. Furthermore, quantitative computer-based measurements are reproducible, as all measuring points on the digital objects can be saved.

This chapter gives a brief introduction to state-of-the-art techniques for 3D data acquisition and computer-based data analysis. Some applications in paleoanthropological research are presented to demonstrate that 3D technology has become an indispensable tool, complementing classical methods in physical anthropology.

27.1 Introduction

Reconstruction plays a crucial part in working with fossils. In a narrow sense, most fossils need to be reconstructed before the scientist can start to elicit information from the specimens, as they tend to be broken or distorted through
diagenetic processes. In a broader sense, not only the fossil itself needs to be reconstructed but also, as an ultimate aim, its place in evolution. Therefore, the function of specific morphological features, in combination with a paleoenvironmental approach, must be determined, in order to understand the role of an organism.

A new aspect of reconstruction has been introduced by recent advances in 3D technology: the digital reconstruction of real objects. With the assistance of various computer-based techniques, such as surface scanning or computed tomography (CT), digital object data can be acquired and processed. What computer-based methods have in common is that they are nondestructive. The handling of the fossil takes place within the computerized environment only, so that no actual damage to the original specimen will occur through physical contact. This aspect alone is of great advantage in comparison to conventional reconstruction and measuring techniques. When reconstructing a fossil digitally, all steps are comprehensible and can be reversed without damaging the specimen. Additionally, quantitative computer-based measurements are reproducible as all measuring points on the digital objects can be saved. Therefore, a comparison between measurements of different researchers should provide better results.

CT additionally reveals internal structures that have previously been inaccessible to scientific examination. Along with data acquisition techniques, new powerful computer-based analytical tools have also been developed. Geometric morphometrics, which enables the quantification of shape, is one of the most promising approaches to investigating fossil morphology. Thereby, 3D technology and computer-based analysis have enormously helped to enhance our knowledge of functional morphology of extant and extinct hominids.

This chapter will provide a brief introduction to state-of-the-art techniques for 3D data acquisition (Figure 27.1) and computer-based data analysis. It will point out the great impact those techniques have on paleoanthropological research. With the assistance of a new methodological inventory, much additional information can now be derived from fossil specimens. And it is crucial to gather as much information as possible from fossils due to their rarity and incompleteness. For these reasons, 3D technology has become an indispensable tool, complementing classical methods in physical anthropology. This is most important in the visualization of internal morphological structures, in morphometric and biomechanical analysis, and in the reconstruction of fragmented fossils. With the assistance of some exemplary applications, the significance of 3D technology for paleoanthropology will be elucidated further.
27.2 Medical imaging techniques

Medical imaging techniques have always been employed in paleontological and paleoanthropological research. Only a few years after X-rays were discovered by W.C. Röntgen in November 1895, this new technique was introduced to paleoanthropology by Gorjanovic-Kramberger (1902, 1906) to study the Krapina Neanderthals. X-ray technology offered the possibility to see the internal structure of rare and valuable specimens in a nondestructive way for the first time.

27.2.1 CT

The development of CT in the 1970s (Hounsfield 1972, 1973) provided 2D cross-sectional images that did not suffer from the superimposition of structures as in conventional radiography (Wind and Zonneveld 1985). This technique was applied to paleontological and paleoanthropological studies in the 1980s (Conroy and Vannier 1984, 1985; Wind 1984; Zonneveld and Wind 1985).

CT is an X-ray–based imaging technology that has in the meantime become a standard tool in noninvasive medical diagnosis and the method of choice for investigating skeletal morphology (see also Lange et al. 1988; Jend and Tödt 1989;
A medical CT scanner consists of an X-ray source which emanates a fan beam that tangentially scans a slice of an object and an opposite array of detectors or a detector ring. The X-ray source rotates around a patient or specimen table while the detectors measure the attenuation of the X-rays from multiple directions in the scan plane. Attenuation is dependent on the density properties of the object and is converted into an electronic signal by the detectors. If measured from different directions, the spatial distribution of the attenuation data of one slice can be calculated (see Kak and Slaney 1999 for the mathematical background). These data are then mapped as a cross-sectional gray-scale image depicting different material densities, or attenuation coefficients, with black representing the lowest and white the highest density. Attenuation coefficients are calculated for each voxel (volume element, see later) from the detector measurements and expressed as CT numbers. The CT number scale is given in Hounsfield units (HU) and is defined by a value of −1000 HU for air, and 0 HU for water (Lange et al. 1988; Jend and Tödt 1989; Hoxter and Schenz 1991; Wegener et al. 1992). Being a digital image, the cross-sectional image of a single CT scan consists of a limited number of pixels (picture elements). Each pixel actually represents a voxel as the image depicts a slice of a given thickness. The pixel is allocated the gray value of the average attenuation coefficient in the voxel. The spatial resolution in the plane of the scan is determined by X-ray beam geometry and pixel size. Pixel size depends on the scanner’s pixel matrix and the field of view and can at best be approximately 0.3–0.5 mm in current medical scanners. Perpendicular to the plane, the spatial resolution is significantly poorer as it is determined by the minimum slice thickness, which is currently about 1.0–1.5 mm in high-resolution medical CT (Spoor et al. 2000a). A stack of cross-sectional images is obtained by moving the table so that CT images of contiguous or overlapping slices can be made. In this way, the whole volume of an anatomical feature can be depicted.

A much higher spatial resolution and thinner slice thickness can be obtained by micro-CT scanners. Some of those scanners work comparably to medical CT scanners. Others use an image-intensifier and a frame grabber with video camera to record radiographs in a great number of directions and calculate a 3D volume of CT numbers therefrom. These scanners were originally developed for industrial purposes and not for scanning living organisms. Therefore, they usually have a stationary X-ray source with a rotating object. The spatial resolution and slice thickness can be between 1 and 200 μm, and it is possible to obtain isometric voxels, i.e., identical pixel size and slice thickness (Spoor et al. 2000a). The scan time
with micro-CT is higher compared to medical CT scanners, as indeed are the costs for those custom-built systems. The results, however, especially in studying fossil morphology and small anatomical features compensate for these drawbacks.

Drawbacks occur in CT image quality when scanning fossils that may be difficult to overcome (see also Zonneveld and Wind 1985; Spoor and Zonneveld 1994; Spoor et al. 2000a; Zonneveld 2002). Problems arise because fossils are often highly mineralized and may be filled as well as surrounded with dense sedimentary matrix. If the density of the specimen is too great, or if its dimensions are too large, the X-rays cannot pass through sufficiently. This lack of detector signal results in “frozen noise” artifacts in the image, which are streaks of noise that obscure details. This problem often arises with matrix-filled, highly mineralized crania. To overcome this difficulty the signal must be increased, which can be achieved by increasing the slice thickness. This, however, reduces spatial resolution at the same time, and a best compromise must be found.

Medical CT scanners usually cover a CT number scale that is suitable for showing structures in a living organism. Fossilized bone and sedimentary matrix can have a density that is much higher than the scale’s maximum. Their attenuation coefficients are then assigned the highest possible CT number and corresponding gray value, which is white (white overflow). These overflow artifacts in images are white areas with no visible structural details. Older scanners do not assign the highest CT number to all attenuation coefficients above a maximum, but the CT number drops to the lowest value which is depicted as black after the maximum has been reached and rises up again from there (black overflow). In both cases, overflow artifacts make structural details in the affected area invisible. Additionally, the outline of the structure is altered and volume as well as surface measurements will yield incorrect results. To overcome overflow artifacts, the CT number scale of the CT scanner must be changed so that it covers the highest density of the fossil (Zonneveld and Wind 1985). Most scanners already have an extended CT number scale, while in others recalibration can solve the problem (Spoor and Zonneveld 1994).

Artifacts through “beam hardening” are another drawback when scanning fossils. Beam hardening is the absorption of low-energy X-rays from the spectrum while passing through the specimen, so that the remaining beam becomes proportionally richer in high-energy photons (Kak and Slaney 1999). This occurs within all CT scanners and they are calibrated to compensate for beam hardening. As they are usually calibrated for scanning living organisms, however, more beam hardening may occur in scanning fossils. Recalibrating the CT scanner with, e.g., aluminum for the compression of the CT number scale is a solution (Zonneveld and Wind 1985; Spoor and Zonneveld 1994).
27.2.2 MRI

Magnetic resonance imaging (MRI) was also developed and applied to diagnostic imaging in the 1970s and 1980s (Lauterbur 1973; Mansfield et al. 1976; Mansfield and Pykett 1978; Edelstein et al. 1980; Steiner and Fiegler 1989). Its use in paleoanthropology is limited because it images soft tissue best and is therefore most applicable to specimens *in vivo*. It can, however, provide valuable information for actualistic studies concerning growth and development and is therefore considered briefly here.

MRI, like CT, produces cross-sectional images. However, this technique is not based on X-ray technology but rather maps the relative abundance of protons in the presence of a strong, static magnetic field. Using radio-frequency pulses, the protons are brought from equilibrium into a higher energy state. After the pulses stop, the protons begin to change back to their lower energy state, while emitting energy in doing so. This energy is received as a signal. The strength of the signal is proportional to proton concentration and is further characterized by the chemical and physical environment of the tissue (see Mansfield and Morris 1982; Steiner and Fiegler 1989; Kak and Slaney 1999; Spoor et al. 2000a, Vlaardingerbroek and den Boer 2003; Huettel et al. 2004 for further technical details).

As in CT, the MR signal is calculated for each voxel and its associated pixel and is depicted on a computer screen using a gray scale. Spatial resolution of medical MRI is about 0.7–1.0 mm with slice thickness being 1.0–3.0 mm. High-resolution MRI also exists with a spatial resolution of 156–300 μm and slice thickness of about 300–600 μm (Spoor et al. 2000a). Imaging times in high-resolution MRI are much longer, though, at about 24 h per specimen, so that artifacts in the images caused by movements can occur. Imaging contiguous or overlapping slices of a specific structure allows the mapping of the whole volume. As previously mentioned, hard tissue like mineralized bone, which is proton deficient, produces only a very weak signal and can therefore not be imaged properly. The surrounding soft tissues, like the brain or cartilage, however, emit a strong echo. Ossified tissue can nonetheless be visualized by MRI, as it is silhouetted against the depicted soft tissue.

27.2.3 3D Visualization

After having obtained a stack of contiguous 2D cross-sectional CT or MRI images, a digital 3D reconstruction of the specimen or parts thereof can be conducted using computer graphics techniques. If serial CT sections are examined
individually, even the experienced investigator may have problems in appreciating the 3D morphology of a region of interest. The first attempts at 3D imaging based on a volume of CT slices occurred in the early 1980s (Vannier et al. 1985; Conroy and Vannier 1986; Vannier and Conroy 1989a, b; Hemmy et al. 1994). Today, this technique is commonly applied (e.g., Zollikofer et al. 1995, 1998; Rae and Koppe 2000, 2002; Gantt et al. 2003; Kono 2004).

The most common technique for reconstructing selected structures (e.g., the skull) in 3D is surface rendering. The first step, the extraction of selected tissues (e.g., bone) from the dataset, is called segmentation. This is mostly conducted by thresholding: choosing the range of density values of the selected tissue to be displayed and removing other CT numbers from the image. This procedure has to be carried out for each cross-sectional image. Thresholding sometimes does not provide satisfactory results, especially when, e.g., bone and adhering matrix have similar CT numbers and are therefore difficult to distinguish. Local differences in mineralization of fossil bone add to thresholding problems. Another difficulty is the definition of the threshold as CT numbers at the borderline between tissues (e.g., bone and air) may differ from the “typical” values due to the averaging of attenuation coefficients in each voxel/pixel (partial volume averaging) (Spoor et al. 2000a, b). If the chosen threshold range is too small, parts at the outline of the chosen structure may be removed. If it is too large, the neighboring tissue will not be totally removed. Interactive segmentation in which the scientist is able to mark regions of interest manually can improve segmentation results. Edge detection and region growing software tools may also help (Spoor et al. 2000a). In the case of fossilized bone and adhering matrix, it is sometimes still impossible for the experienced observer to distinguish these two tissues, even visually. Prossinger et al. (2003) introduced a filter that helps in distinguishing bone from matrix and enables segmentation. Further progress will hopefully facilitate this very time-consuming task of segmentation.

After removing nonrelevant tissue from the slices, the images are combined along the border-lines of selected structures so that the structure surface can be viewed three-dimensionally. Subsequently, a surface model can be rendered by triangulation of the segmented object. Sequentially stacked slices sometimes result in uneven surfaces of the digital model. Smoothing algorithms can be applied to enhance the realistic appearance of the object. Computer graphics techniques in general have improved significantly, resulting in increasingly realistic images. The accuracy of the digital reconstructions, however, is limited by the spatial resolution in the scan plane, the slice thickness, and the slice increment (Spoor et al. 2000b).

Shadowing the 3D images by virtual light sources produces the effect of spatial depth and facilitates examination of the digital model. Viewing the
internal structures of the digital reconstruction of the specimen is enabled by merely cutting away or making translucent overlying structures of the model.

27.2.4 Medical imaging techniques in paleoanthropology

One application of CT data has already been demonstrated: the computer-based preparation of fossils. Fossils may be filled with or enclosed in sedimentary matrix so that physical preparation might destroy the precious fossil or at least parts thereof. Internal structures cannot easily be accessed nondestructively, or the position of the fossil parts within the stony matrix cannot be estimated reliably. In a different scenario, a fossil may have already been physically prepared and also reassembled previously and is now fixed with plaster and glue, but the reconstruction is debatable. In all these cases, it might be desirable to isolate the fossil parts and to reconstruct the specimen (anew) without damaging the fossil. Through digital preparation, matrix and fillings as well as formerly applied artificial reconstruction material can be removed from the digital bone. As described previously, matrix and other material can be removed by thresholding during the segmentation process (Vannier et al. 1985; Conroy and Vannier 1986). Manually guided “electronic chisels” (Ponce de León 2002), semiautomatic software tools (Spoor et al. 2000a), or filter algorithms (Prossinger et al. 2003) can refine the preparation results. Some specimens that are too fragile to be physically prepared, or are filled with matrix that cannot be accessed without damaging the specimen, can now be investigated digitally (Spoor and Zonneveld 1999; Prossinger et al. 2003).

A good example for a fossil specimen that has been digitally dis- and reassembled is Le Moustier 1. The adolescent Neanderthal skull was physically reconstructed several times, the last time in 1925 by Weinert, and was partly disassembled during the late 1960s (Thompson and Illehrhaus 1998). Different workers have applied electronic techniques to reconstruct the skull digitally from serial CT scans (Thompson and Illehrhaus 1998; Ponce de León and Zollikofer 1999; Ponce de León 2002; Thompson et al. 2002). Plaster, glue, and other matrix, which was hard to differentiate from the fossil material by mere vision, could be removed from CT images. All bony fragments were able to be isolated and reassembled digitally (Figure 27.2), and it was then possible to investigate the specimen again in the light of current knowledge (Ullrich 2005).

Computer-based reconstruction of specimens like Le Moustier 1 may also be necessary because the fossil fragments are distorted by the taphonomic process they underwent (see Grupe Volume 1 Chapter 7). A physical reconstruction of a
fossil from deformed pieces must remain unsatisfactory as measurements that are taken on it in the end yield incorrect results. Strong bilateral asymmetries, e.g., indicate plastic distortions. Sometimes, CT analysis reveals internal deformations that have remained undetected previously as in the Steinheim cranium (Prossinger et al. 2003). Digital reconstruction can be very useful in solving this problem by computerized deformation of the fossil back to its original shape. Vannier et al. (1985) already mentioned the method of deforming skeletal parts of the computer model to visualize pathological pre- and postoperative conditions, and suggested it for digitally reconstructing deformed fossil skulls. In the meantime, the methods for plastic distortion of digital objects have improved enormously. Information about the in situ situation of the fossil is needed. The original shape can be obtained by applying reverse forces to the gravitational

Figure 27.2
Reconstruction of LeMoustier 1 from micro-CT scans. Fragments are differentially shaded. It should be noted that only the original bony parts were used for this reconstruction. Micro-CT scanning was done by B. Illerhaus at the BAM, Berlin, image processing and reconstruction were conducted by J.L. Thompson using the software AVS (courtesy of B. Illerhaus)
forces that caused the deformation (Zollikofer et al. 1998; Ponce de León and Zollikofer 1999; Ponce de León 2002; Thompson et al. 2002; Zollikofer 2002; Zollikofer and Ponce de León 2002). However, as not every diagenetic event that led to postmortem modifications can be fully assessed, the resulting shape still remains tentative, and the resulting reconstructions are sometimes hotly debated.

Taphonomic processes not only result in plastic deformation but also in fragmentation of the fossil and loss of skeletal parts. Fractured objects can be reconstructed digitally by joining the fragments in their original orientation. Assembling the pieces is usually a stepwise procedure and demands practical knowledge of skeletal anatomy. If teeth are present, occlusional relations can be used to match antagonistic teeth and bony fragments of upper and lower jaws (Thompson and Illerhaus 1998; Ponce de León 2002). The bony labyrinth serves as an “anatomical compass” to orient temporal and adjacent bones (Zollikofer et al. 1995, 1998). Other fragments are added and realigned piece by piece to allow the edges of the bone fragments to meet correctly (Thompson and Illerhaus 1998).

Missing parts can be completed by either mirroring existing pieces of the same specimen (Zollikofer and Ponce de León 1995) or by adding corresponding parts from similar specimens that can be scaled and deformed (Kalvin et al. 1995; Zollikofer et al. 1998; Weber et al. 2003; Neubauer et al. 2004).

In summary, one can say that the reconstruction of fossil specimens (in a narrow sense) has gained enormously by applying computer-based methods such as digital plastic deformation and mirroring. Another advantage of reconstructing a fossil digitally is that all steps are comprehensible and, most important, can be reversed anytime without damaging the specimen. Once a best solution for the reconstruction is found, the original specimen can be reconstructed physically, either by using the original parts or by means of rapid prototyping (see Section 27.4).

The fossil cannot only be reconstructed digitally. Computer-based analysis also provides a direct dataset for established as well as new means of investigating the specimen.

Determining the volume of the endocranial cavity from a stack of CT scans was one of the first applications the new technology offered which was used (Vannier et al. 1985; Conroy and Vannier 1986), and it is still one of the main concerns of CT investigations (Conroy et al. 1998, 2000; Weber et al. 1998; Recheis et al. 1999; Falk et al. 2000; Thompson et al. 2002; Neubauer et al. 2004). Cranial capacity, representing brain size, is measured as an important item of taxonomic classification. Measurement of other cranial cavities, like sinuses, to better understand their functional and phylogenetic significance is also possible nondestructively by using CT scans (Prossinger et al. 2000; Rae and Koppe 2000, 2002; Prossinger et al. 2003). Determination of the volume of hollow features can
either be carried out on single cross-sectional CT images that are measured individually or directly on the digital 3D reconstruction. For measuring the volume, in the former case, the 2D area of the region of interest in the image is measured and multiplied by slice thickness, and the measurements for all slices finally summed up (Conroy and Vannier 1985; Rae and Koppe 2000). To determine the volume in the latter case, an algorithm counts the number of voxels and multiplies by voxel size (Prossinger et al. 2003). To obtain a best estimate for brain size, some refined techniques can be applied to this simple calculation (Weber et al. 1998).

With the development of high resolution and micro-CT scanners, very small internal anatomical features are now possible to visualize. One of the most famous examples for this kind of application is the investigation of the bony labyrinth. This is located inside the petrous part of the temporal bone and is thereby only accessible nondestructively by applying CT. It houses the hearing organ as well as the sense organs for the perception of movement and spatial orientation. Zonneveld and Wind (1985) showed that it is possible to image the labyrinth of fossil hominids with a series of high-resolution CT scans. They hoped to be able to obtain information about the physical properties of hearing and speech of fossil species by reconstructing the middle and inner ear. At the same time, they realized that the orientation of the semicircular canals might be used as an indicator of locomotor behavior (Wind and Zonneveld 1985). Spoor and colleagues followed this approach to determine the emergence of bipedalism in hominid evolution (Spoor and Zonneveld 1994, 1998; Spoor et al. 1994, 1996). They found that canal size and locomotor behavior were correlated and humans had larger anterior and posterior semicircular canals and a smaller lateral canal than other extant primates (Spoor and Zonneveld 1998). Among fossil hominids, the australopithecines show apelike proportions, while *Homo erectus* is the earliest species to have humanlike semicircular canal morphology. This is interpreted as being indicative of habitual bipedalism in *H. erectus* and facultative bipedalism in australopithecines. The similar proportions of semicircular canals in the great apes are considered the ancestral hominid condition so that labyrinthine morphology can provide information about locomotor behavior as well as phylogeny in hominid evolution. The Neanderthal bony labyrinth has a derived morphology that allows taxonomic specimen identification (Hublin et al. 1996; Spoor et al. 2003) (Figure 27.3). This is also possible for subadult individuals as the labyrinth already possesses its adult morphology at the time of birth (Spoor and Zonneveld 1998).

A different application in which the small size and fossilized nature of objects make special demands on the method of investigation is the determination of enamel thickness. Enamel thickness as an important functional and taxonomic
character has been the center of many investigations since the 1970s. Most studies have used sectioned teeth (Gantt 1976; Molnar and Gantt 1977; Martin 1985; Shellis et al. 1998; Ulhaas et al. 1999; Smith et al. 2003), which has considerably limited the number of specimens investigated, especially of fossils. Conventional radiography (Sperber 1985; Molnar et al. 1993; Smith and Zilberman 1994) as well as medical CT (Conroy 1991; Grine 1991; Macho and Thackeray 1992) have been employed to measure enamel thickness nondestructively with controversial results. Only high-resolution CT yielded satisfactory results (Spoor et al. 1993; Schwartz et al. 1998). The extremely high density of dental enamel and the small size of hominid teeth make it difficult to get exact measurements. Overflow artifacts (see Section 27.2.1) give the enamel cap a swollen appearance, which results in overestimation. Even in early high-resolution CT, with slice thickness of 1.5 mm and using overlapping scans, exact measurements below a critical distance were not possible due to still insufficient spatial resolution (Spoor et al. 1993). Current micro-CT scanners that are capable of a slice thickness of between 50 and 80 microns allow exact measurements of enamel thickness (Gantt et al. 2003; Kono 2004; Olejniczak and Grine in press) (Figure 27.4). Additionally, not only linear and area measurements in a fixed plane but also enamel distribution and volume can be measured using 3D reconstructions of the tooth crown. This technology therefore at last allows the nondestructive investigation of enamel thickness of fossil teeth. New insights into functional and phylogenetic implications of enamel thickness can also be expected by employing new 3D parameters. Due to the noninvasiveness and the good visualization of small features by high-resolution and micro-CT, other dental characters can now be studied as well, e.g., tooth root morphology (Kupczik et al. 2003).
MRI, as mentioned before, has a limited use for paleoanthropology as it can mostly visualize soft tissue. However, MRI is the method of choice for ontogenetic analyses (Bookstein et al. 2003) or studies of brain development with respect to phylogeny (Semendeferi et al. 1997; Semendeferi and Damasio 2000; Jeffery and Spoor 2002).

A different application of MRI in paleoanthropology is its use in soft-tissue reconstruction. The MRI image of a human head contains information about the depth of soft tissue that covers the skull. For facial reconstructions, the digital image of muscles, fat, skin, etc. overlying the facial skeleton instead of statistical
information of tissue depths can be used to reconstruct a face over a given skull. This kind of reconstruction has no immediate scientific value. But it is a means for getting realistic soft-tissue reconstructions of fossil hominids for museum exhibitions or schools.

### 27.3 Surface data

Many computer-based applications in paleoanthropology currently use digital data that have been acquired by CT scanning. Even though CT imaging offers much information about the internal morphology of the structures of interest, the data acquisition process is expensive in every regard. It can be long and tedious, the data volume that has to be stored is large, and the costs are relatively high, given the possibility to use a good CT scanner. If only the surface of an object is to be measured, however, several means for surface data acquisition exist (Dean 1996; Petrov et al. 1998; Beraldin et al. 2000), which serve the purpose well or even better than surface reconstructions from serial CT scans.

#### 27.3.1 Surface scanning

Several techniques have been developed to capture 3D surface information. These divide into optical and nonoptical, contact and noncontact, and active and passive sensing methods (Breuckmann 1993; Petrov et al. 1998; Beraldin et al. 2000; Godin et al. 2002; Zhang et al. 2002). Optical noncontact techniques ensure damage free data acquisition. While passive techniques operate under ambient lighting conditions and use natural features on the surface of an object to reconstruct it in 3D, active methods create features on the surface by controlled projection of light. Therefore, active techniques can be applied to a broader range of objects with fewer restrictions to surface texture and lighting. For scanning small and midsize objects and short measurement ranges up to a few meters, active optical triangulation systems are the most appropriate. Meanwhile, these systems provide photorealistic representations of shape and texture with reasonable speed (Figure 27.5).

Optical triangulation systems either use laser light or pattern projection. Triangulation scanners generally consist of a projector that either projects a laser beam or a grid pattern onto the object and one or more digital cameras. The light is projected from a known position in a known direction. In laser scanning systems, the direction of the returning light is measured by an angle sensor. With a fixed distance between projector and angle sensor, the known projection
angle, and the measured angle of the returning light, the position of the illuminated point can be calculated with regard to the system, according to the rules of triangulation. In systems using pattern projection, camera and projector are arranged in a fixed distance and angle to each other. The camera captures grid patterns on a target surface. Many patterns are known as “codes” (e.g., Gray code) and are meant to uniquely express light directions. If the surface is flat, the lines of the pattern appear straight and undeformed. Any surface features result in changes of the projected pattern. The lines will appear curved, bent, or otherwise deformed. Knowing the projection angle of the pattern, triangulation formulas as well as other mathematical techniques can be applied to calculate the profile of the surface (Beraldin et al. 2000; Godin et al. 2002; Zhang et al. 2002).

To obtain data on the whole object, the object has to be scanned in different views. To capture the whole surface, there is either a rotating specimen table or the light source rotates around the object. In some mobile systems, the scanner has to be manually moved around the object. Each view results in a 3D point

Figure 27.5
Reconstruction of H. rudolfensis mandible UR 501 from Malawi from surface scans. Fossil specimen is originally fragmented in three pieces. Fragments were scanned with a Breuckmann optoTop® scanner. Image processing and reconstruction was done with software Polyworks® (courtesy of F. Schrenk)
cloud. To visualize the complete surface, point clouds of the different views have to be matched and merged. Finally, a 3D point-cloud results that represents the complete surface of the object. Computer graphic techniques can again be used for surface rendering.

Color texture reconstruction of the digital model can be achieved by different methods. Color images of the object can be mapped on the 3D model surface using perspective projection. The quality of the result depends on the color image quality and the integration of images from different views that usually show slight differences in color and lighting. For laser scanning approaches, one solution is the use of a polychromatic laser so that color information of the illuminated spot can be measured directly. However, this technique requires scanning in very high resolution and is therefore not appropriate for scanning very large objects (Blais et al. 2003).

Optical-scanning systems have some limitations, though. Because the camera has to discern the light that is projected on the object, very dark or reflective surfaces and transparent media make it difficult to get good scanning results. Deep cavities or surfaces that are covered by other structures result in missing values as the projected light does not reach these object areas. Coating the specimen surface with matt and light powder, using different lighting conditions and scanning the object from various angles, however, enables the complete recording of most specimens.

One major advantage of laser-scanning systems is that the laser light can be easily filtered out from the ambient illumination so that they can be operated in daylight conditions. Pattern projection systems rather need dark surroundings as the grid pattern cannot be captured unambiguously by the camera in too bright lighting conditions.

Accuracy and resolution of both pattern projection and laser systems are reasonable. The main drawback in using laser light is speckle noise. Speckle is produced when a rough surface is illuminated with coherent laser light. Speckle appears as grains attached to a surface and results in measurement uncertainty. However, meanwhile special filters can reduce speckle noise and thereby enhance accuracy up to 0.01 mm. Resolution of laser optical triangulation is limited by the diffraction of the laser light. Usually, a resolution of 0.05–0.01 mm can be achieved. In pattern projection, using one grid code alone limits measurement resolution considerably. However, combining a code pattern with a phase-shift approach results in theoretically infinitely high resolution. In practice, resolution is limited by electronic noise and errors due to quantization of gray levels in the intensity image (Beraldin et al. 2000) and is comparable to laser-scanning systems. However, accuracy and resolution actually always depend on the object surface and texture.
27.3.2 Surface scanning in paleoanthropology

Surface scanning has been introduced to paleoanthropological studies only recently. Laser scanning was employed by Aiello et al. (1998) and Wood et al. (1998) for the investigation of joint surfaces of fossil limb bones. The aim of these studies was to determine if isolated fossil bones could be attributed to one individual or at least one species. These authors, as well as Harcourt-Smith et al. (2004), who applied geometric morphometric methods (Section 27.3.4), were able to show that the congruence of reciprocal joint surfaces can be ascertained on the taxonomic level and that refined statistical methods will enable the identification of unassociated skeletal elements of individuals.

Small objects like teeth can also be digitized by surface scanning. Ungar and colleagues (Ungar and Williamson 2000; M’Kirera and Ungar 2003; Ungar and M’Kirera 2003; Ungar et al. 2002) recorded hominoid molar surfaces with the help of a laser scanner and applied geographic information systems (GIS) to model, characterize, and compare tooth surfaces as landscapes. In a different approach, Kullmer et al. (2002, 2004) and Ulhaas et al. (in press) used pattern projection scanning to obtain digital models of hominid and other primate teeth. Both approaches aim at a topographic analysis of occlusal relief. The application of 3D methods for the analysis of tooth surfaces allows for the inclusion of new measurements that quantitatively record morphological traits that could be described only qualitatively before. Molar surfaces are measured and compared with regard to such features as surface area, basin volume, occlusal relief, and cusp slopes. Established measurements like tooth length can also be taken on the digital models and allow comparison with earlier investigations. The most important aspect of these 3D approaches is that it is now possible to include worn molars into analyses (Ungar and M’Kirera 2003; Ulhaas et al. 2004, in press). Many worn fossil molars that have never been used so far for functional and/or systematic analyses can now be studied with the assistance of a new methodological inventory.

Last but not the least, it is also possible to reconstruct fragmented fossils using 3D surface scans of the fragments, as can be seen on the example of the *H. rudolfensis* mandible UR 501 (Figure 27.5). As the use of surface-scanning techniques is quite new in paleoanthropology, many applications still wait to be explored.

27.3.3 Digitizing landmarks

Digitizers are systems for measuring 2D or 3D coordinates of landmark locations. Landmarks are homologous measuring points that can be reliably and repeatedly
located. The coordinate data can be recorded in different ways. First of all, a fixed coordinate system has to be defined in which the object is located. Therefore, the origin of the coordinate system and the endpoints of at least two axes are recorded. A digitizer usually consists of a stylus, which is used like a pen to mark the measuring points. The stylus can be part of a mechanical arm with several rotating joints. The positions of the joints allow the calculation of the landmark coordinates. Otherwise, 3D data can be recorded using interacting electromagnetic fields, the transmitter, and the receiver field. Each field is determined by a Cartesian coordinate system, formed by three antenna coils which are at right angles to one another. Whenever the transmitter coils are pulsed, a current is induced in each receiver antenna coil. The strength of the current is dependent on both the distance and the relative orientation of the transmitter and receiver coils. The resulting nine current values are used to calculate the 3D position of the landmark (Foley et al. 1990). Other systems use ultrasound waves to record the position of the stylus, or markers to mark the measuring points, while sensors or cameras determine the planes they lie in and thereby ascertain the 3D coordinates of the points (Foley et al. 1990). These latter systems are currently not used for anthropological applications. Although the electromagnetic system is especially prone to inaccuracies due to interferences by electric or metallic objects, such as the computer screen, the precision of measuring landmark data in anthropological applications is assessed as satisfactory (Baab et al. 2003; Marcus et al. 1997). Measurement accuracy of about 0.15 mm for the electromagnetic and 0.25 mm for the mechanical digitizer, however, is not sufficient to measure such objects as small primate teeth (Ungar and Williamson 2000).

27.3.4 Landmark data and geometric morphometrics

Landmark points can be joined linearly to portray the object surface. The closer the points are together, the finer will the surface representation be. Surface details though cannot be depicted and therefore cannot be studied using this method. Landmark data permit statistical analysis of biological form using either traditional morphometric methods (Sokal and Rohlf 1995) or the advantages of geometric morphometrics (Bookstein 1991; O’Higgins and Jones 1998; O’Higgins 2000; Richtsmeier et al. 2002). “Form” contains information about size and shape. While “size” is usually defined as a specific length, area, volume, or centroid and can be measured quantitatively, “shape” usually remains immeasurable and can only be described qualitatively. In geometric morphometrics,
variation in shape, defined as the spatial distribution of landmarks, can be analyzed independent of size by applying a set of morphometric tools. Among those are superimposition methods like Procrustes superposition (Rohlf and Slice 1990; Bookstein 1991), deformation methods like thin-plate splines (Bookstein 1991), and analyzing tools like Euclidean distance matrix analysis (EDMA) (Lele and Richtsmeier 1991), principal component analysis (PCA) (O’Higgins and Jones 1998), and relative warp analysis (Bookstein 1991). For further information about geometric morphometric methods, the reader is referred to Bookstein (1991); Bookstein et al. (1999, 2003); O’Higgins (2000); O’Higgins and Jones (1998); Richtsmeier et al. (2002); Rohlf and Slice (1990).

Landmark data can either be recorded on the original specimens using a digitizer or on the digital objects obtained by CT, MRI, or surface scanning. CT and MRI also offer the possibility of measuring internal landmarks. Anthropology has a long history of measuring biological form. Therefore, many applications exist that have used the opportunity to include new 3D landmarks and the methods of geometric morphometrics in the study of size and shape variation.

Harvati (2002, 2003a, b) investigated the morphological variation in cranial landmark configurations between Neanderthals and anatomically modern Homo. She showed that the morphological distance between these two groups is so large that the hypothesis that Neanderthals represent a subspecies of *H. sapiens* cannot be supported. Delson et al. (2001) used a comparative morphometric analysis based on landmarks to determine the systematic position of the Sabungmacan 3 *H. erectus*. Bookstein et al. (1999) were able to show substantial external cranial differences, but stable inner frontal bone morphology, between early and modern Homo.

Comparing not only different taxa but also different ontogenetic stages using geometric morphometrics allows a better understanding of growth patterns. Bookstein et al. (2003) and Mitteroecker et al. (2004a, b) were able to separate phylogenetic and ontogenetic processes that have an impact on cranial shape in great apes and humans. A comparison between Neanderthal and anatomically modern human immature crania (Ponce de León and Zollikofer 2001; Zollikofer and Ponce de León 2002) showed that morphological differences between these two groups appear very early during ontogeny, possibly prenatally, and are maintained during later development. The researchers conclude that morphological differences therefore have a genetic rather than an environmental background and that Neanderthals and modern humans should be separated at the species level.
27.4 Rapid prototyping

It was shown that many different analyses can be conducted on 3D digital models of fossil specimens. Nonetheless, is there sometimes a need for having a physical model to work with. It is easier to comprehend the spatial distribution of anatomical features when holding the specimen in one’s hands and turning it around, than if one is restricted to looking at a 2D image of a 3D model on the computer screen.

A means of getting the digitally reconstructed fossil into reality is rapid prototyping. Rapid prototyping technologies were originally developed for reverse engineering to get 3D hardcopies of digitally designed objects. The best-known technology is stereolithography (Ghezal and Stucki 1992; Bresenham et al. 1993; Zollikofer and Ponce de León 1995). With stereolithography, a 3D object is constructed in layers by computer-guided polymerization. For this purpose, the 3D digital model has to be converted into a special file format that allows cross-sectioning of the object into slices. Inside a container with photosensitive polymers, the slices are built consecutively by laser curing. A laser beam is guided by the computer across a liquid layer, according to the shape of a cross-section. Where the laser beam hits the liquid, the photopolymers harden. After one layer is completed, an elevator inside the container moves downward to submerge the object and to cover the hardened layer with a new layer of liquid of defined thickness, usually between 0.1 and 0.5 mm. The process continues until the object is completed. To prevent the stereolithographic model from sagging during the construction process, small supporting struts are added into the object that can be removed after the final hardening of the complete model. The accuracy of stereolithographic models is about 0.15 mm, while most of the error is due to inaccuracies from CT scanning and data processing.

The first stereolithographic model that was used for anthropological studies was the replica of the skull of the Tyrolean Ice Man (Seidler et al. 1992). The mummy was CT scanned and CT data used for the plastic reconstruction of the skull. The skull was built in two parts so that inner cranial details were directly visible. Access to internal parts as well as transparency of the reconstruction material enable direct visual comparison of traits like endocranial sinuses (Seidler et al. 1997). Zollikofer et al. (1995) used stereolithographic models to visually ascertain the correct alignment of anatomical elements during the digital reconstruction process. Stereolithography is also an alternative to casting fragile fossils. The specimen has to be surface-, MRI-, or CT scanned only once, which is a nondestructive procedure, as shown earlier. The 3D copy of a specimen obtained by stereolithography is as accurate as a conventional cast. Furthermore, multiple copies can be done without any loss of accuracy, as occurs with traditional
methods (Zollikofer et al. 1998). Additionally, just parts of the model that allow insight into the internal anatomy, or enlarged replicas of very small units, can be produced with this technique.

27.5 Conclusions and future prospects

It is clear that computer-based paleoanthropology has made much progress during the last two decades. Meanwhile, computer-based 3D technology has become a powerful means for paleoanthropological investigations. Data acquisition and analysis can be carried out nondestructively within a computerized environment. Fragile fossils or specimens that are embedded in sedimentary matrix and fragmented or deformed by taphonomic processes can be reconstructed with the new technologies. In reconstruction and analysis, all steps are reproducible and can be reversed without damaging the original specimen. With the assistance of a new methodological inventory, much additional information can now be derived from fossil specimens. The more information can be obtained, the better the reconstruction not only of the fossil itself but also our comprehension of its place in evolution.

Many fossils wait to be analyzed with the new techniques, so the amount of data that can be used in comparative studies will grow. Mobile systems are already available that allow surface scanners or high-resolution CT scanners to be brought to the fossils. This will further ensure a growing number of scanned specimens.

An increase in power and affordability of personal computers and the emergence of more appropriate software will convince more researchers to apply the new technologies. Imaging quality of surface scanners and in MRI is increasing, as well as commercial availability of micro-CT scanners. Furthermore, one of the disadvantages of the early systems has been overcome: universal formats that allow data exchange between different systems have been developed. Therefore, the call for setting up digital data archives and sharing data is not surprising (Spoor et al. 2000a; Weber 2001, 2002). Additionally, the combination of different data acquisition techniques, such as CT and surface scanning or CT and MRI, in multimodal applications is possible and will enhance analysis efficiency.

In future applications, not only static analysis but also studies of dynamic processes, like locomotion and mastication, will be conducted. The new kind of data and new measuring tools are not meant to replace traditional methods in physical anthropology but rather to complement them. Computer-based reconstruction technology has already become an indispensable tool in
paleoanthropological applications and will hopefully become a standard procedure in future investigations.

Acknowledgments

I would like to thank all of my colleagues who always tried to answer my questions concerning the amazing techniques available today. Special thanks to Bernhard Illerhaus, Anthony Olejniczak, Friedemann Schrenk, and Fred Spoor for providing figures to illustrate some applications.

References


Hounsfeld GN (1972) A method of and apparatus for examination of a body by radiation such as x-ray or gamma radiation. Patent specification 1283915. The Patent Office


Jend HH, Tödt HC (1989) Arbeitsbuch Computertomographie. Schnetztor/Byk Gulden Pharmazeutikon, Konstanz
O’Higgins P, Jones N (1998) Facial growth in Cercocebus torquatus: an application of...

Olejniczak AJ, Grine FE (in press) High-resolution measurement of Neandertal tooth enamel thickness by micro-focal computed tomography (μCT). S Afr J Sci


Rae TC, Koppe T (2002) 3D imaging and measurement in studies of cranial pneumatization. BAR Int Ser 1049: 11–16


Zonneveld F (2002) Applications and pitfalls of CT-based 3-D imaging of hominoid fossils. BAR Inter Ser 1049: 5–9
28 Prospects and Pitfalls

Jean-Jacques Hublin

Abstract

Paleoanthropology is primarily rooted in the study of fossils and the analysis of sites. Dependence on these resources leads to challenges resulting from the difficulty in gaining access to scarce, precious, and sometimes overprotected materials, and from issues of control over field sites. The development of virtual paleoanthropology can sometimes be a way to partially solve the first problem. However on some occasions, the access to and utilization of numerical data has also become an issue of dispute. In parallel, recent advances in studies focusing on microstructures, isotopic composition, and paleogenetics require direct sampling of the fossils. The trend in paleoanthropology is to integrate approaches from different scientific fields, and this is especially visible in developmental sciences, genetics, and environmental studies. In the meantime, dealing with human evolution remains a sensitive topic, subject to clear ideological and religious biases. The interest of the media and of the public in this science does not always contribute to an objective approach to the questions. Finally, among other issues, the expansion of paleoanthropology studies in developing countries must contend with the decline of a colonial mode of thinking.

28.1 Introduction

In many respects, paleoanthropology is a paradoxical science. Although it addresses the oldest origins of humans, the discipline itself developed quite recently in the history of science. The first fossil hominid specimen on record, an immature Neandertal skull, was discovered in 1830 in Engis (Belgium). However, it was not until the end of the nineteenth century that certain fossil specimens were truly accepted by the majority of the scientific community as evidence for the evolutionary process that gave birth to our species. Although very significant discoveries occurred during the first half of the twentieth century and provided the basic framework for paleoanthropological studies, the last three decades of that century witnessed a spectacular increase in the available fossil record as well as in major advances in the knowledge of past environments and the chronological background of human evolution. Moreover, the birth of paleogenetics added
a new dimension to the analysis of relationships among fossil hominid species. The current state of the discipline results not only from methodological progress but also from a major effort of field research. For the public, the media, and students, it is always a matter of amazement and sometimes of criticism to realize that the field of paleoanthropology is such a changing terrain. If prediction, in science as in many other domains, is always a difficult exercise, it is even more challenging in this particular science, which is relatively newly born and still rapidly evolving.

Another distinctive aspect of paleoanthropology lies in the fact that the study of fossil specimens remains the core of the discipline. These are rare and precious objects. It is often emphasized that for the study of some extinct groups the specimens are fewer than the specialists who analyze them, and sometimes the competition is harsh. After an undefined period during which they remain under the relatively rigid control of individuals or groups responsible for publishing descriptions of them, they are usually curated in museums or other institutions. While the specimens are in their possession, the curating institution may also restrict access to the fossil specimens, emphasizing their conservation rather than their scientific study. On the one hand, there is therefore strong pressure to consult the fossil material, and on the other, growing restrictions that result from a multitude of reasons, ranging from the desire to maintain scientific monopoly to the legitimate policy of protection of fragile and valuable objects. In parallel, new techniques for the study of the specimens have been developed. However, these new methods sometimes resolve issues and sometimes generate new difficulties.

### 28.2 Virtual paleoanthropology

Since the 1980s, new techniques in medical and industrial imaging have revolutionized the fields of human paleontology and physical anthropology (Wind 1984; Wind and Zonneveld 1985; Zonneveld and Wind 1985) allowing the development of what has become commonly called as “virtual paleoanthropology.” The growing use of computed tomography as well as industrial imaging techniques (microtomography and laser scanners) has allowed the production of 3D images of fossil specimens. Combined with stereolithography and other techniques of 3D printing, these virtual representations have opened a number of new possibilities for the analysis of the specimens. Most notable among these are

- Virtual extraction and reconstruction (including correction of plastic distortions) (Kalvin et al. 1992; Zollikofer et al. 1995; Ponce de Leon and Zollikofer 1999)

• Precise quantitative analysis of inaccessible internal structure (including tiny structures, such as middle and inner ears, bony tables, vascular foramina, etc.) and their comparison with living references (Zonneveld et al. 1989; Hublin et al. 1996; Spoor et al. 1996)
• 3D morphometric analysis with the development of new mathematical tools (Harvati 2002)
• Modeling of ontogenetic processes, biomechanical properties, and of evolutionary changes themselves (Ponce de León and Zollikofer 2001; Mitteroecker et al. 2004)

Growing evidence suggests that, with increasing frequency in the future, the anatomy of fossil hominins will be systematically studied not from the specimens themselves but from virtual representations. Principal among all the new possibilities opened by virtual paleoanthropology is the reduced need to manipulate the real objects. Consequently, these techniques should be welcomed by many curators. However, they also raise new questions. One is related to the quality of the data. So far, the CT scanners that have been used to acquire the data are primarily those available in the medical environment. Although they evolve rapidly, machines of this type have their own limitations and are not specifically designed to explore fossilized specimens filled by dense sediments. The resolution of the 3D pictures produced in this way is far from satisfactory in many cases. Their quality could be greatly improved by the implementation of high-resolution industrial scanners designed for paleoanthropological use. Such machines have already been acquired by a handful of institutions in the USA, Japan, and Europe.

Among the pitfalls related to the development of virtual paleoanthropology is a shift from a situation where access to the fossil specimens was difficult to a situation where access to the numerical data is even more challenging. Curators are sometimes reluctant to allow repeated acquisitions of these numerical data, while the techniques and equipment evolve rapidly. Often, the data are monopolized by those who acquired them initially, and they are hard to upgrade. In the long term, databases may develop in some institutions and on the Internet. To date, however, the development of such databases has faced insuperable difficulties. The commercialization of some of these data by the institutions concerned or the simple trading of data between teams will remain, for sometime, the only alternative.

28.3 Into the matter of bone

In a somewhat opposite direction to virtual paleoanthropology, there are a number of other new approaches that have been developed in the fields of
human paleontology, physical anthropology, and archeology. Such techniques were initially based on rather invasive analyses of the specimens, inherited primarily from histology and geochemistry. However, with the rise of non- (or less) destructive methods, this field is rapidly expanding. In the future, study of the actual fossil remains will likely be reserved for the kinds of analyses that cannot be performed on virtual representations. At present, such analyses include, on the one hand, histological approaches mainly addressing bone and tooth microstructures and, on the other hand, chemical analyses addressing either geochronological or paleobiological questions.

Microstructural studies have developed mostly in the field of dental anthropology. The recognition of different types of incremental mineralized structures in the dentine and the enamel since the middle of the twentieth century has led to systematic analyses of their variation in extant and fossil primates (Dean 1987; Stringer et al. 1990; Lieberman 1993; Zhao et al. 2000; Dean et al. 2001; Martin et al. 2003; Schwartz et al. 2003; Smith et al. 2003). This development has been made possible by the improvement of technical equipment such as the scanning electron microscope, the confocal microscope, and computer-assisted microscopy for 3D visualization. The interest in these studies comes from the knowledge that microstructures could be the main, if not the only way, to assess life history in extinct species (FitzGerald 1998). This issue has been given increasing attention in an evolutionary perspective since the genetic bases of development have become better understood, and their importance for evolutionary changes better appreciated. Future research in this direction will certainly include extensive work on modern variability and more experiments to assess the biological significance of accretional microstructures and their relevance for calibrating the growth patterns of extinct species. Although it is possible to work on externally visible features, such as perikymata (Ramirez Rozzi and Bermudez de Castro 2004), a drawback of these methods is the necessity of slicing precious fossil specimens to analyze fine internal microstructures. However, the technique of thin slicing has greatly improved, and it is possible today to “rebuild” a specimen after analysis following minimal destruction of tissue. In the future, new techniques of imaging may also partly resolve this problem. Although, to date, it remains a very expensive technique, the use of synchrotron allows access to bone and tooth microstructures without destruction (Tafforeau et al. in press).

Chemical analyses of fossil specimens have been aimed at reconstructing paleobiological features and are mostly concerned with the extraction of organic molecules. Nonorganic chemical properties of the fossil remains are primarily relevant to the determination of their geological age and are marginally useful in addressing paleobiological issues. To date, DNA and collagen have been the main targets of the research on ancient biomolecules. Techniques based on the use of
Restriction enzymes have allowed the duplication and subsequent sequencing of tiny and rare fragments of DNA chains. So far, this work has been based primarily on the analysis of mitochondrial DNA, which is smaller and much more abundant than nuclear DNA. The sequencing of a fragment of mitochondrial DNA of the Feldhofer 1 (Neandertal) specimen in 1997 opened a new era of paleoanthropological studies (Krings et al. 1997). Future development of this research will involve the reconstruction of the entire sequence of the mitochondrial DNA in specimens such as Neandertals. With the development of new techniques, future work will also address the issue of nuclear DNA in fossil hominids. However, the natural degradation of DNA under given physical conditions imposes a chronological barrier that today seems oddly unsurpassable. Another serious problem in paleogenetic studies comes from the potential for contamination. Paradoxically, the DNA of modern and relatively recent humans remains very difficult to identify as genuine fossil DNA and to distinguish from subsequent contamination (Serre et al. 2004). Studies on the taphonomical processes affecting the deterioration of DNA chains in the archeological or geological deposits may provide an answer to this problem.

Isotopic composition of the mineral portion of hominid fossils has been used to assess biological issues. These studies face the difficult questions of the taphonomic transformation of the chemical composition of fossils in the geological layers (Radosevich 1993; Fabig and Herrmann 2002). Most researchers have thus focused on the more stable component of bone, the protein collagen (Schoeninger 1985; Ambrose 1986; Bocherens et al. 1991; Bocherens et al. 1997; Richards et al. 2000; Richards et al. 2001). Collagen has been the primary source for the analysis of stable isotopes such as oxygen, nitrogen, and carbon. These isotopes are fixed in the living tissues antemortem, either at an early stage of individual development (in the teeth, for example) or at some time before death. They are an essential source of information about the environment and the diet of individuals during their lifetimes. One constraint of these studies is that they are limited by the long-term preservation of collagen. One can expect that, as with the study of recent archeological series, such analyses in the future will bring unexpected knowledge on issues such as migration, seasonality, or even mating strategies among relatively ancient hominids. New research into extracting other longer surviving proteins, such as osteocalcin, has the potential to provide material for isotopic studies for much older material. An interesting development comes from the combination of microstructural studies and isotopic analyses to assess fine timescale changes in the diet or the environments of fossil individuals (Humphrey et al. 2004). The extension of isotopic analyses to new elements may also lead to interesting developments in this field. For example, sulfur isotopes in collagen along with strontium and oxygen in minerals can tell us...
about migration and movement patterns. Isotopic studies (especially of oxygen) will also likely contribute to a much greater knowledge of the environmental conditions and their rapid variation in the continental environment, a topic that so far remains much less explored than in oceanic environments and the ice caps.

28.4 Understanding evolutionary processes

The reconstruction of the evolutionary history of hominoids, and more specifically of fossil humans, has for a long time focused primarily on taxonomic and phylogenetic questions. Important methodological progress has been made in this field during recent decades. In particular, the use of cladistic approaches has provided a better theoretical background. Although these approaches also have their own limitations (Trinkaus 1990), they became indispensable for assessing the significance and the polarity of features. However, it should be underlined that the development of mathematical techniques to analyze size and shape, including 3D morphometrics, has led to the regression of some studies to a precladistic stage. The emphasis placed on the shape distances should not lead us to forget that morphological similarity is not a reliable way to analyze phylogenetic relationships when the polarity of the features is not taken into account.

A major problem, discussed extensively in recent years, centers around the features used by paleoanthropologists for cladistic analyses. These discussions have focused on features’ significance and on their relationships either to genetic determinism, or to environmental conditions and behavior, or to an interaction between the two. Beyond these discussions lie issues such as the independence of such features in their development and their homology when one passes from one species to another (Lieberman 1999; Wood 1999). These are critical questions for the analysis of the fossil evidence and the reconstruction of phylogenetic relationships from a parsimonious perspective. However, one may be reasonably optimistic in this matter, as experimental data and a better understanding of the precise genetic and epigenetic mechanisms underlying the development of features will resolve these questions. These may also bring answers to related problems such as the discrepancy sometimes underlined between biomolecular evidence and phylogenetic reconstructions based on the analysis of morphological features of the phenotype (Collard and Wood 2000; Strait and Grine 2004). They may also bring new light to the debate surrounding modular versus integrative models in the biological development of extinct organisms (Wagner 1996; Wagner and Altenberg 1996; Williams and Nagy 2001; Winther 2001).
As far as the recent stages of human evolution are concerned, it is reasonable
to expect that taxonomic and phylogenetic issues may become of minor interest
in the future, as the main taxa are identified and their phylogenetic relationships
understood. However, better understanding of the variability not only in extinct
taxa but also in living forms remains a crucial issue. Although some have
predicted the decline of such anatomical studies, it is still striking to contemplate
our lack of knowledge of the variability in living humans with respect to many
features commonly used in paleoanthropological research. After several centuries
of anatomical studies, we still desperately need valid longitudinal data on the
growth and development of many anatomical features. This lack of data is even
more dramatic when one considers the populations of living apes, our closest
relatives in the animal world; most of them will likely become extinct in the wild
before they have been properly studied.

Research may focus more on paleobiological issues. Changes in growth and
development processes during life history, in terms of timing and pattern, are
increasingly seen as powerful mechanisms to explain evolutionary changes.
Studies of extinct species consider this dimension with increasing frequency,
and developmental trajectories will hopefully be identified for different taxa as
will the effects of epigenetic phenomena. This is, of course, dependent on an
increase in the available paleontological material and also on a greater interaction
between paleoanthropology and developmental genetics. 3D morphometrics and
other mathematical tools have been identified recently as powerful tools for the
reconstruction of those developmental trajectories that can sometimes be mod-
eled (Ponce de León and Zollikofer 2001). Establishing reliable tools to assess
the calendar ages of immature individuals in the fossil record is of crucial
importance in this matter (Coqueugniot et al. 2004), and developing studies of
skeletal microstructure seem an inevitable way to address this problem.

Other aspects of the biology of extinct species might also become accessible
through progress in the extraction of biomolecules such as proteins. Osteocalcin
has recently been extracted from Neandertal remains and sequenced (Nielsen-
Marsh et al. 2005). In the future, extraction of proteins or lipids from ancient
material may even shed new light on the physiology of our ancestors and cousin
species. It should also be noted that the extraction of ancient proteins and their
sequencing may be extended much further back in time as some of these
molecules seem to resist taphonomic degradation much better than DNA.

In recent years, paleodemographic questions have become more and more
interesting to paleoanthropologists. Topics such as life history and longevity
are critically important to understanding the biological and social adaptations
of ancient groups, as well as to addressing questions of learning time during
individual life and the transmission of knowledge from one individual to another. Other paleodemographic parameters that appear to be important are the questions of population densities in given areas in a geological timeframe, and their possible catastrophic variation in relation to environmental changes. Although paleodemographic parameters have long appeared unreachable (Bocquet-Appel and Masset 1982, 1996), different methods of evaluating size fluctuation in ancient populations have emerged from genetic or paleogenetic studies. Although paleogenetic studies on the Neandertals did not revolutionize our views of their phylogenetic relationships, they did bring a new way to assess genetic variability in ancient populations, and consequently, size changes (Currat and Excoffier 2004; Serre et al. 2004). The animal models appear to be a tempting alternative for testing demographic fluctuations and their effects on genetic variability during the recent periods of the Pleistocene (Orlando et al. 2002). Such knowledge will allow a better understanding of the possible effects of demographic crashes and genetic bottlenecking on evolutionary processes themselves as well as the relative roles of genetic drift and natural selection. Combining isotopic analysis and microstructural studies will also allow the garnering of information such as the weaning age of fossil individuals. A fine knowledge of climatic environmental changes, sometimes perceptible on the scale of one human life, also brings new light to the way human populations have adapted biologically and culturally. In this perspective, a better integration of biological and cultural evidence seems necessary for a more thorough understanding of human evolution.

### 28.5 Chronology

The determination of the geological age of the fossil hominids is central to paleoanthropological work. Until recently, such determination has centered on the application of radiometric methods to the archeological context of the discoveries. Although available methods require improvement in their precision and in their calibration, their direct application to hominid specimens also represents major progress, especially for specimens anciently discovered and/or for which the context is unknown or inaccessible. Once more, the development of such approaches has been limited initially by the destructive aspects of these investigations. However, the emergence of new techniques, such as laser ablation, makes the analysis of light or heavy isotopes on precious specimens almost completely nondestructive. In the future, these studies may be applied routinely and become the best way to establish a precise chronology.

As far as the radiometric methods based on C\textsuperscript{14} disintegration are concerned, the development of mass spectrometry has allowed the direct dating of fossil
specimens by requiring only small amounts of matter for analysis. In the coming years, one may expect the establishment of a better calibration curve between 30,000–40,000 BP that will make the method more reliable in this time range of crucial importance to the history of modern human dispersal (Bard et al. 2004). In addition to calibration, contamination remains a major problem and the origin of the sampled carbon must be securely established; methods such as the ultrapurification of the collagen may allow for the control of this factor. Another future development will be to work with organic carbon from biomolecules that can be identified by sequencing as genuine fossil molecules belonging to the extinct organisms from which they were extracted. In practice, all this means that many C<sup>14</sup> dates acquired during the last decades may become meaningless because of the limitations of the techniques used to establish them. Large databases that have been built to process these dates by thousands and that provide a picture of biological and cultural evolution of humans, especially, at the time of the replacement of Neandertals by modern humans in Europe can be improved by a critical assessment of the compiled dates. However, the screening process has provided very contrasting pictures and has not satisfactorily solved all questions. This probably results from some bias in the selection process of the data retained, partly depending on the views that different authors have on the evolutionary and peopling processes involved. Eventually, such databases may become obsolete and will need to be rebuilt with more reliable geochronological information.

### 28.6 Picture of an ancestor

Every human society has built up physicotheological explanations to deal with the question of its origins. In the historical record, such explanations have been furnished by religions and mythologies, but from the middle of the nineteenth century, western societies substituted a scientifically based explication model for biblically based explanations. The question of human origins became the concern of the scientists rather than priests. However in 1863, T. Huxley, a main supporter of Darwin’s views, wrote: “The question of questions for mankind, the problem which underlies all others, and is more deeply interesting than any other, is the ascertainment of the place which man occupies in nature, and of his relationship to the universe of things. Whence our race has come; what are the limits of our power over nature; to what goal are we tending are the problems which present themselves anew, and with undiminished interest, to every man born into the world.” Almost 150 years later, this issue is still not free of ideological, if not metaphysical, constraints. This may partly explain why the public has developed
such an interest in this field of science. Aside from the inherent attraction of pictures of extinct worlds, any piece of evidence in paleoanthropological studies also becomes a matter of opinion and feeling, even for nonknowledgeable audiences. Prehistory is also consistently the topic of novels, films, and documentaries in which science always has to defend itself against fantasy. Museums and educational centers have become incredibly successful with this subject, showing increasing sophistication in their ways of responding to the demands of the public for pictures or 3D reconstructions. Although some of these reconstructions are today produced (Figure 28.1) using advanced techniques, they have their own limitations. It is certainly possible to provide reasonable reconstructions of the general anatomy of well-known fossil species. However, to date many fine anatomical details, such as skin, hair, and eyes, are beyond the range of

Figure 28.1
A reconstruction of a pre-Neandertal at the Landesmuseum für Vorgeschichte (Halle, Germany). Accurate methods have been developed to reconstruct soft tissues in fossil hominids. However today, like in the past, the picture of ancient humans primarily remains a projection of our own fantasies
scientific assessment. Unfortunately, they are also of crucial importance to the
way kinds of humans other than our own species appear to us. “Scientifically
based” reconstructions of fossil hominids filling the museums in Europe and
America may say much more about our way of perceiving human diversity than
about the actual aspect of these hominins. In this respect, the progress of our
reconstructions since the beginning of the twentieth century may be more limited
than is often assumed. Underlying notions of humans as belonging to different
species, and possibly contemporary in the past, are difficult to integrate not only
for the public but also for scientists from sister disciplines. The humanist
framework within which the human and social sciences developed in our uni-
versities may explain the difficulty that cultural anthropologists and archeolo-
gists, attached to the notion of “uniqueness” of the human being, face in dealing
with notions such as ape cultures or the multispecific nature of hominins. More
generally, in the post–Second World War era, new conceptions understandably
developed around human diversity that provided an ideological framework to
which paleoanthropological evolutionary models had to adapt. The questions of
Neandertal nature and abilities and their relationship to extant humans, in other
words the last well-documented divergence in our phylogenetic tree, is one arena
in which science and ideological preconceptions have clashed in a complex way.

In this interaction between scientists and the public, the media play an
important role. There are many reasons why scientists need to communicate
with the public. One is that the interest of the public partly justifies society’s
investment in this field of pure research. Another may result from more personal
reasons. For a department, for a team, for an individual, the visibility of obtained
scientific results becomes increasingly important as it impacts on possible per-
sonal promotion and political decisions to support this field of research in general
or a project in particular. Thus, the scientist and the reporter face each other in a
dialogue where each needs the other. The reporter needs material for exciting
articles; the scientist needs a reporter for publicity. In the past few decades, this
interaction has become increasingly important and has sometimes led to unde-
sirable effects. One obvious pitfall is that the public and reporters are more
interested in some issues than in others. Those problems most debated in the
media may be of limited interest to the scientific community and vice versa. Sex
between Neandertals and modern humans is an example of a question universally
addressed to the paleoanthropologists, and one danger is that, in the need to be
well represented in the media, scientists might be led to responding to questions
or even developing research interests geared to public attention. It is amazing to
see how well-developed press services have become not just in institutions
dedicated to public education and communication but also within research
structures as well.
Recently these interactions between scientists and the media have entered a new dimension, as personal issues or rivalries between individuals have become of themselves a matter of interest for the press. High-profile international scientific journals have developed “people sections” that deal almost exclusively with these subjects. Reporters have therefore become part of the scientific debate and actors in rivalries, by promoting opinions and sometimes by fueling controversies and conflicts in a way designed to make their articles more exciting.

28.7 Unbalanced ecology of paleoanthropology

Paleoanthropology is based on the study of specimens to which access, as we have seen, may be difficult. It is also based on sites and fieldwork. The result is that it can be a highly territorial activity, to an extent unequaled in other fields of science. Indeed, aside from scientific problems, the paleoanthropologist must also face a series of political and ethical issues. Although specimens discovered and published a long time ago should be fully accessible to the scientific community, this is not always the case. The situation is even more complex regarding specimens soon after their discovery and/or after a partial description. So far, the scientific community has not established a consensus on resolving such questions. The discoverer of a new specimen has a scientific and moral right to it and is granted priority in providing a scientific analysis of this material, alone or in collaboration with other specialists. However, this well-accepted notion is often blurred by complications. One such complication can result from the multiplicity of the discoverers involved and from lack of agreement at the time of the discovery or later. And many discoveries occur during the course of archeological excavations conducted by teams that were not anticipating the possibility of hominin discoveries. Another problem can result from an abusive extension of the time spent between a discovery and the publication of a reasonably comprehensive description of it. The competition that is natural in science is sometimes displayed in a negative form, such as preventing challengers from accessing material, which may be used as a more efficient way to surpass them than producing better scientific results.

Similar situations have developed with respect to site and field access. Most commonly, research teams obtain the monopoly of the study of one site or a geographical area for a certain length of time in order to conduct a defined scientific program. However, apart from this formal arrangement, there are a number of situations where informally and based on political influence, tradition, or nationalistic issues, institutions manage to secure a geographical domain of influence. This is often the case when excavations or field operations are conducted by scientists from western countries in less developed areas. Often,
this is facilitated by the fact that the studied areas are located in countries lacking indigenous research in the field of paleoanthropology and/or scientists trained in this discipline. In such cases, the work is conducted under the authorization of local administrations that are primarily preoccupied by the conservation of their national patrimony. Here again, different institutions or scientific communities may develop some level of competition in guaranteeing their access to the field. Different countries have developed different regulations restricting the exploitation of archeological and paleontological material. Situations in which fossil specimens can simply be transferred from the country of their discovery to scientific institutions or museums in Europe or America have almost completely disappeared today.

Many scientists have felt compelled to develop balanced collaboration with the host countries in which they conduct their research, in particular, by helping with the conservation of the material and with the training of local scientists. Furthermore, many countries outside of the western world have managed to develop their own programs and scholars, and the trend is more and more to develop joint projects, although a very unequal equilibrium in terms of financial contribution and/or scientific expertise can lead to bitter conflicts. One thing is clear: the time of scientific colonialism is over (or will be over soon), and although in an ideal world, fossils and sites should be accessible to “everyone,” we are still far from this situation. Archeological and paleontological materials are considered as part of the national patrimony in most countries, yet the future of the field lies in fair and fruitful international collaborations.

References

Huxley TH (1863) Evidence as to man's place in nature. Williams and Norgate, London
1 Primate Origins and Supraordinal Relationships: Morphological Evidence

Mary T. Silcox · Eric J. Sargis · Jonathan I. Bloch · Doug M. Boyer

Abstract

There are five major scenarios that have been advanced to account for the early events in the origination of the order Primates: a transition from terrestriality to arboreality, the adoption of a grasp‐leaping mode of locomotion, the evolution of features for visual predation, an adaptation to terminal branch feeding occurring during angiosperm diversification, or a combination involving terminal branch feeding followed by visual predation. These hypotheses are assessed using both neontological and fossil data. Of the five scenarios, the angiosperm diversification hypothesis is not contradicted by modern data and is found to be the most consistent with the fossil record. In particular, the evolution of features for manual grasping and dental processing of fruit in the earliest primates (primitive plesiadapiforms), and the subsequent development of features for better grasping and more intense frugivory in the common ancestor of Euprimates and Plesiadapoidea, is consistent with a close relationship between early primate and angiosperm evolution. All the other scenarios are less consistent with the pattern of trait acquisition through time observed in the fossil record. Consideration of non‐euprimates (e.g., scandentians and plesiadapiforms) is found to be essential to viewing primate origins as an evolutionary process rather than as an event.

1.1 Introduction: what is a primate?

Perhaps the most fundamental issue facing students of primate origins can be summarized by a simple question: what is a primate? A clear concept of the diagnosis and taxonomic composition of Primates is essential to providing a coherent understanding of when and why the order separated from the rest of Mammalia. Attempts to define the order Primates have typically started by considering which features of modern primates are present in multiple primate species and are distinctive relative to other mammals. Four major adaptive
complexes of traits have been recognized as characteristic of primates of modern aspect (=Euprimates Hoffstetter 1977; see Mivart 1873; Le Gros Clark 1959; Napier and Napier 1967; Martin 1968, 1986, 1990; Szalay 1968; Cartmill 1972, 1992; Szalay et al. 1987):

(1) Traits associated with grasping. These include relatively longer hand and foot phalanges, a divergent thumb and big toe, and digits tipped with nails rather than claws.

(2) Traits associated with leaping. Although such features have been lost in some extant primates (e.g., *Homo sapiens*), the most primitive euprimates have leaping characteristics that include hindlimbs that are long relative to the forelimbs and modified ankle bones.

(3) Traits associated with improvements to the visual system. These features include large eyes, convergent orbits, and a postorbital bar or septum. The larger and more complex brain of modern primates compared to other euarchontans (see below) may also be associated, in part, with this complex. A smaller apparatus for the sense of smell is presumably associated with an increasing reliance on vision as well, leading to a short snout and proportionally reduced related areas of the brain.

(4) Dental traits associated with herbivory. Relative to specialized insectivores, primates possess teeth that are low crowned, with blunt and bulbous (bunodont) cusps and broad talonid basins, which are features related to eating non-leafy plant materials (e.g., fruit) rather than insects or meat.

Presence of a petrosal bulla has also often been cited as an ordinally diagnostic primate trait (e.g., Cartmill 1972) but its adaptive significance (if any) is unclear.

Defining the order Primates using observations on living taxa as a starting point is problematic because these traits are unlikely to have evolved simultaneously. Instead, a definition that recognizes the process of primate evolution and that encompasses the earliest, possibly stem, members of the order will have greater explanatory power. We will return to this problem below.

### 1.2 Ecological scenarios for primate and euprimate origins

Researchers investigating primate origins have typically focused on building an ecological scenario that could explain the evolution of one or more of these adaptive complexes. The earliest such scenario is the arboreal hypothesis of
primate origins, which traces its roots back to the work of G. Elliot Smith and F. Wood Jones in the early part of the twentieth century. The arboreal hypothesis was extended and broadly popularized by W.E. Le Gros Clark (1959). In this hypothesis, grasping extremities were seen as having value for more secure climbing, and the distinctive primate orbital features were explained as being useful for judging distances in the trees during leaping.

All of the other ecological scenarios that have been developed assume a life in the trees for ancestral primates but seek to go beyond simple arboreality to consider more specific types of behavior. Szalay and colleagues (Szalay and Delson 1979; Szalay and Dagosto 1980, 1988; Szalay et al. 1987; Dagosto 1988) considered a derived locomotor mode, grasp-leaping, to have driven the evolution of most of the features that characterize euprimates, including those of the visual apparatus. They linked the ability to rapidly jump from branch to branch with the need to be “... subsequently securely anchored” (Szalay and Delson 1979 p 561) to the landing point. Visual changes were relevant to judging distances in rapid, leaping locomotion (Szalay and Dagosto 1980). In this hypothesis, anatomical changes for grasp-leaping were preceded by a shift to a more herbivorous diet in the primitive primates (i.e., plesiadapiforms) ancestral to Euprimates (Szalay 1968, 1972; Szalay and Dagosto 1980; Szalay et al. 1987).

Cartmill (1972, 1974, 1992) focused on visual predation as key to the origin of Euprimates. The visual predation hypothesis linked visual features beneficial to accurately gauging the distance to prey items with grasping, clawless hands and feet that could provide both a secure hold on narrow supports, and a prehensile apparatus for snatching prey. Because he thought they lacked orbital specializations and grasping features, Cartmill advocated excluding plesiadapiforms from Primates.

Sussman (1991; Sussman and Raven 1978) suggested a link between the origin of Primates and the Cenozoic diversification of angiosperms (i.e., trees that produce fruit and flowers). He agreed with Szalay that a key event in early primate evolution was the invasion of the “arboreal mixed feeding adaptive zones” (Sussman and Raven 1978 p 734) in the Paleocene. This involved increased use of non-leafy plant resources by early primates as angiosperms developed features that made them more tempting to non-insect seed and pollen dispersers, such as specialized flowers and larger fruit. With the appearance of still larger propagative plant organs (e.g., fruit, seeds) near the Paleocene–Eocene boundary, the ancestral Euprimates developed features for entering terminal branches to better exploit these resources.

There are two major classes of data that have been used to assess the relative validity of these various ecological scenarios. The first “tests” various ecological functions assigned to character complexes in the different models using the
comparative method. The second employs the fossil record to document the sequence of anatomical changes that occurred in primate evolution and seeks to tie these changes to adaptive shifts.

### 1.2.1 Comparative method

Cartmill (1970, 1972, 1974) assessed the then prevalent arboreal hypothesis from the point of view of the diversity of modern arboreal animals. He argued that if living in an arboreal habitat could explain the distinctive features of primates, then these traits should also be found in other arboreal forms, and particularly in the arboreal members of groups that also include terrestrial species. Cartmill found that arboreal animals in general do not have features similar to those seen in modern primates. For example, arboreal squirrels are not more primate-like than terrestrial squirrels in grasping traits, such as a reduction of the claws, or in vision-related features like the degree of orbital convergence. Nonetheless, arboreal squirrels are successful at many of the same behaviors practiced by primates, including making reasonably long jumps and foraging among slender branches. He argued that forward facing orbits, while enhancing stereoscopy, decrease parallax and with it the ability to judge distance at longer ranges. For this reason, orbital convergence is not a very useful trait for gauging distances in a jump but is very effective for visualizing objects close to the face. On the basis of these comparisons, it seems unlikely that the distinctive adaptive complex of euprimates can be simply linked to a shift to an arboreal mode of life.

Cartmill’s own hypothesis of primate origins, visual predation, has also been attacked from the standpoint of modern analogy. Garber’s (1980) (but see Crompton 1995) work on the tamarin *Saguinus oedipus* revealed a feeding mode similar to that discussed by Cartmill in an animal that possesses claw-like nails. The evolution of claw-like nails in callitrichines, who nonetheless practice visual predation, undermines the association between this feeding behavior and specialized grasping. Sussman (1991) (see also Crompton 1995) pointed out that most living primates are omnivores, not specialized insectivores, and that their methods of prey capture often emphasize scent and hearing over vision—tarsiers have actually been observed to capture their prey with their eyes closed (Niemitz 1979)! Furthermore, many of the insects eaten by extant primates are crawling forms captured on the ground, not flying forms plucked from the air. If living primates are not typically specialized visual predators, it is not clear why we would expect morphologically similar extinct species to be.
Alternatively, some authors (Sussman and Raven 1978; Sussman 1991; Crompton 1995) have sought analogues for early primates among frugivores, such as old world fruit bats. However, Ravosa and Savakova (2004) recently found no relationship between the degree of orbital convergence and the level of frugivory in a broad array of modern taxa. Nonetheless, the absence of primate-like visual features in some modern visual predators (e.g., mongooses and tupaiid treeshrews; Cartmill 1992) suggests that there is no simple relationship between forward facing orbits and this mechanism for prey capture in mammals. A possible “solution” to this criticism, that the euprimate-like mechanism of orbital convergence for prey capture is only needed in nocturnal animals (Allman 1977; Cartmill 1992), would be refuted if the earliest euprimates were not nocturnal (see below).

Sussman’s (1991; Sussman and Raven 1978) angiosperm diversification hypothesis has been criticized based on the lack of an association between the diversification of angiosperms and the evolution of adaptations in arboreal marsupials that converge on those seen in primates (Cartmill 1992). Nonetheless, in a study of the somewhat primate-like South American didelphid marsupial *Caluromys derbianus*, Rasmussen (1990) did find some support for Sussman’s model, in that a substantial part of its diet comes from terminal branch fruit feeding in a manner similar to modern primates. This study additionally provided some support for Cartmill’s model, in that *Caluromys* is also a visual predator that occasionally grasps prey with its hands. However, *Caluromys* does not have particularly convergent orbits (Rasmussen 1990), weakening substantially the link between this anatomical feature and visual predation. Rasmussen’s study can be seen as providing a fifth composite scenario for primate origins that has ancestral primates initially venturing out onto terminal branches to find fruit and other plant parts, with the secondary evolution of features for prey capture to capitalize on the insect resources they found in this milieu (Rasmussen 1990; Cartmill 1992).

The grasp-leaping scenario of primate origins (Szalay and Delson 1979; Szalay and Dagosto 1980, 1988; Szalay et al. 1987; Dagosto 1988) is less susceptible to criticisms based on modern analogy than the other ecological scenarios because it does not depend on a general ecological relationship for its validity. Rather, it stems from a “fossil-first” approach to considering adaptive change, beginning with the evolutionary transitions documented in the fossil record, and then attempting to determine their adaptive meaning in a form-functional context. This highlights a major contrast in approaches to the question of primate origins between the major combatants in the debate. Under Szalay’s approach, the unique origins, constraints, and evolutionary histories of different mammalian
lineages mean that adaptive explanations applied to one group need not apply to any other. Cartmill, however, argues that “[t]he only evolutionary changes we can hope to explain are . . . parallelisms: recurrent modifications that show up over and over again in different lineages for the same structural or adaptive reasons . . .” (Cartmill 1993 p 226).

One problem with Cartmill’s approach is that it assumes that all adaptive shifts of interest must be parallelisms because otherwise there would be no possibility of explaining them. This reflects a more general problem with the use of modern analogy to “test” hypotheses about evolution. Any historical event is by definition a unique occurrence, even if it is more or less similar to other such unique occurrences that have taken place in other lineages. There is no reason to believe that everything that has happened once has necessarily happened twice. The evolutionary process that produced primates began at a unique starting point (i.e., the divergence of this clade from the rest of Mammalia) and finished at a unique endpoint (i.e., the diversity of extant species). The starting point was heavily constrained by the evolutionary history of what went before, and the adaptive significance of the features evident at the current endpoint is dependent not only on the current usage of a given trait, but also on the biological needs of all the animals that existed along the evolutionary lineage leading to a particular modern species. Modern non-primates that appear similar to primates might have passed through series of adaptive stages quite different from those experienced by our distant ancestors and thus may have arrived at their current form by a very different path. For this reason, arguments that ancient marsupials did not acquire their primate-like traits as a result of angiosperm diversification are not directly relevant to the question of whether or not primates did. The study of modern primates, or modern non-primate analogues, in isolation cannot provide a demonstrably accurate picture of the process of primate origins—it can only yield hypotheses that are more or less plausible for subsequent testing by the fossil record.

A somewhat analogous situation occurred in the early history of human paleontology. In the early part of the twentieth century, quite plausible scenarios were proposed that suggested either a large brain or bipedal locomotion as being the first-occurring distinct human trait (Lewin 1987). With the discovery of the australopiths—primitive human ancestors with adaptations for bipedal locomotion but relatively small brains (McHenry and Coffing 2000)—any “brains first” scenario was decisively falsified, no matter how plausible it may have seemed on the surface. And so too must any ecological scenario of primate origins be considered falsified if the predicted pattern of trait acquisition is not matched by the fossil record.
1.2.2 Fossil record

1.2.2.1 Primate supraordinal relationships

The origin of Primates was a process, not a single event. The distinctive primate traits listed in the introduction likely did not arise in an evolutionary instant. If this is true, the elucidation of the adaptive process of primate origins relies centrally on knowing the relationships of taxa at the base of the primate tree to understand the evolutionary steps taken to build the first euprimate.

In recent years, the supraordinal relationships of primates among living mammals have been sought in Archonta, a group named by Gregory (1910) to include primates, dermopterans (flying lemurs or colugos), treeshrews, chiropterans (bats), and elephant shrews. Modern conceptions of Archonta (McKenna 1975; Szalay 1977; Novacek and Wyss 1986; Novacek 1992; McKenna and Bell 1997) exclude elephant shrews but postulate a close evolutionary relationship between the first four groups based on morphological similarities (Silcox et al. 2005). Within Archonta, primates have traditionally been allied with treeshrews, a group that was for many years included in the order Primates (Carlsson 1922; Le Gros Clark 1925, 1926, 1959), but which are now classified in their own order, Scandentia (Butler 1972; McKenna and Bell 1997). Novacek (1992) provided a consensus morphological solution for archontan relationships, with links between Scandentia and Primates on the one hand and Dermoptera and Chiroptera (=Volitantia) on the other.

More recent molecular and morphological studies have questioned this view. Molecular studies have consistently failed to find a close relationship between bats and the other archontans—rather, chiropterans have generally grouped with carnivores and ungulates (Pumo et al. 1998; Miyamoto et al. 2000; Liu et al. 2001; Murphy et al. 2001a, b; Springer et al. 2003, 2004). The other three orders have clustered together in many recent molecular studies (Adkins and Honeycutt 1991; Waddell et al. 1999; Liu et al. 2001; Murphy et al. 2001a, b; Springer et al. 2003, 2004), leading Waddell et al. (1999) to propose the name Euarchonta for a group including Primates, Dermoptera, and Scandentia. Within Euarchonta, consistent molecular support for the traditional grouping of primates and treeshrews has also been lacking, with a Scandentia-Dermoptera (=Sundatheria; Olson et al. 2005) link being supported by several recent analyses (Liu et al. 2001; Murphy et al. 2001a, b; Springer et al. 2003, 2004). If this hypothesis is correct, it implies that the best model for the ancestor of primates based on extant forms is represented by the reconstructed common ancestor of Euarchonta, not by treeshrews in isolation. In particular tupaiids, the diurnal family of treeshrews most often used for comparison to primates (Beard 1993a), become less relevant as
ancestral primate models. The sole living member of the family Ptilocercidae, *Ptilocercus lowii*, is the extant treeshrew closest to the base of Scandentia (Olson et al. 2004, 2005) and shares many more features than tupaiids with dermopterans (Sargis 2001a, 2002a, b, c, d, 2004, in press). These shared features are present in *Ptilocercus* and dermopterans in spite of some fundamental differences between their locomotor modes (gliding in dermopterans, arboreal quadrupedalism in *Ptilocercus*; Sargis 2001a, 2002a, b, c, d, 2004, in press), implying that they may be ancestral for the common ancestor of Sundatheria, or even of Euarchonta. As such, *P. lowii* might provide us with the best living model for the common ancestor of Primates (Figure 1.1).

Discovery of new fossil specimens has also led to significant challenges to the view of archontan relationships portrayed by Novacek (1992). In 1990, two papers (Beard 1990; Kay et al. 1990) documented specimens of taxa classified in the plesiadapiform family Paromomyidae that were interpreted as providing strong support for a plesiadapiform–dermopteran relationship. Beard (1989,
1990, 1993a, b) even went so far as to postulate that paromomyids and another plesiadapiform family, Micromomyidae, practiced a dermopteran‐like, mitten‐gliding mode of locomotion.

The relevance of these views to primate origins is related to the systematic position of plesiadapiforms. “Plesiadapiformes” is a paraphyletic grouping of extinct fossil mammals known from the Paleocene and Eocene of North America, Europe, Asia, and possibly Africa (Russell 1964; Beard and Wang 1995; Fu et al. 2002; Smith et al. 2004; Tabuce et al. 2004; Silcox and Gunnell in press). Represented by more than 120 species classified into 11 or 12 families (depending on whether or not the enigmatic African azibiids are included: Silcox 2001; Tabuce et al. 2004), plesiadapiforms represent a very diverse radiation and form a significant component of the faunal record from many Paleocene localities (Rose 1981; Gunnell et al. 1995). The systematic position of plesiadapiforms has been a long‐standing matter of debate. Most early workers classified plesiadapiforms in Primates, often specifically in Tarsiidae (Matthew and Granger 1921; Gidley 1923), based largely on dental similarities. More recent workers have often viewed plesiadapiforms as the first radiation of the order Primates, more primitive than any modern group (Szalay and Delson 1979; MacPhee et al. 1983; Szalay et al. 1987). This hypothesized plesiadapiform–euprimate relationship is not without its detractors. Martin (1968) and Cartmill (1972) were the first to seriously question the primate status of plesiadapiforms. In both of these cases, each worker’s emphasis was on forming a clearer definition of the order Primates, rather than providing an outline of primate supraordinal relationships, leaving plesiadapiforms either in limbo or classified as part of a wastebasket “Insectivora.” Such an approach ignores the possibility that identifying ancestral fossil forms would be essential to understanding the adaptive steps leading to the common ancestor of Euprimates.

Wible and Covert (1987) also suggested removing plesiadapiforms from the order Primates on the grounds that cranial evidence was more supportive of a scandentian–euprimate tie than a plesiadapiform–euprimate one. They argued that the dental evidence linking plesiadapiforms to euprimates consisted only of ill‐defined “trends” (Wible and Covert 1987 p 9). This conclusion was not based on any detailed consideration of teeth, however, which is particularly problematic since the euprimate–plesiadapiform relationship had always been supported largely by dental evidence. Furthermore, one of the critical basicranial features that they cited in support of a Scandentia–Euprimates clade excluding plesiadapiforms has since been found in a paromomyid plesiadapiform (a bony tube for the internal carotid nerves and/or artery; Silcox 2003).

Beard’s (1989, 1990, 1993a, b) and Kay et al.’s (1990, 1992) suggestion of a plesiadapiform–dermopteran relationship offered a novel reason for excluding
plesiadapiforms from Primates. Although their papers agreed on the plesiadapiform–dermopteran tie, the analyses performed by these authors produced hypotheses of relationships for the rest of Archonta, including the placement of Euprimates, which were otherwise almost entirely contradictory (\textodot\ Figure 1.2).

Figure 1.2
(a) Kay et al.’s (1992: fig. 11) maximum parsimony cladogram, redrawn with “Euprimates” substituted for their “Primates” following the terminology in use here; (b) relationships of taxa included by Kay et al. taken from Beard’s (1993a: fig. 10.1) analysis. Beard did not explicitly include any lipotyphlan insectivores, implying that he considered them to lie outside of the ingroup. The only areas of congruence between these hypotheses are the monophyly of Chiroptera and the relationship between the paromomyid Ignacius and the dermopteran Cynocephalus. Otherwise they are entirely in conflict.

Kay et al. (1992) failed to support Archonta or Euarchonta, with plesiadapiforms and dermopterans falling outside of lipotyphlan insectivores on their cladogram, and presumably far distant from bats, primates, and tree shrews (\textodot\ Figure 1.2a). Likely because of their common focus on cranial data, Kay et al. (1992) agreed with Wible and Covert (1987) that Scandentia is the sister taxon to Euprimates. Beard (1993a; \textodot\ Figure 1.2b), on the other hand, found a close relationship between modern dermopterans and plesiadapiforms (his “Dermoptera”; Beard 1993a, b) and euprimates, a group he named Primatomorpha. The sister group to this clade in Beard’s (1993a) analysis was Chiroptera, with Scandentia falling out as the basal-most group of archontans. These two analyses suggested very different taxa as being central to issues of primate origins—plesiadapiforms and dermopterans on the part of Beard and tree shrews on the part of Kay and colleagues.
Beard’s (1989, 1990, 1993a, b) and Kay et al.’s (1990, 1992) conclusions have been challenged by numerous studies on both phylogenetic and functional grounds (Krause 1991; Szalay and Lucas 1993, 1996; Wible 1993; Wible and Martin 1993; Van Valen 1994; Runestad and Ruff 1995; Stafford and Thorton 1998; Hamrick et al. 1999; Stafford and Szalay 2000; Boyer et al. 2001; Bloch and Silcox 2001, 2006; Silcox 2001, 2003; Bloch and Boyer 2002, 2003; Sargis 2002d, in press; Silcox et al. 2005). Some of the shortcomings identified by these authors include a lack of appropriate sampling of Scandentia (Sargis 2002d, in press), problems with the identifications of specimens (Krause 1991; Hamrick et al. 1999; Boyer et al. 2001), features that were missed or misinterpreted (Simmons 1994; Stafford and Thorton 1998; Bloch and Silcox 2001; Silcox 2001, 2003; Sargis 2002d), and a lack of appropriate character sampling in their almost entire exclusion of dental data and omission of hypothesized volitant synapomorphies (Silcox 2001; Sargis 2002d, in press; Silcox et al. 2005). Since the publication of Kay et al. (1992) and Beard (1993a, b), numerous new specimens of plesiadapiforms have been uncovered, documenting previously poorly known or totally unknown anatomical regions (Boyer et al. 2001; Bloch and Silcox 2001, 2003, 2006; Bloch and Boyer 2002, 2003; Figure 1.3). The impact of these specimens also needs to be considered in terms of their relevance to plesiadapiform and euprimate relationships and taxonomy.

Silcox (2001) sought to remedy some of the problems identified above with more extensive character and taxon sampling, including a large dental dataset, inclusion of hypothesized volitant synapomorphies, reconsideration of all characters, and codings for cranial and dental features of the scandentian *P. lowii*. She examined 181 dental, postcranial, and cranial characters for 85 species. Her results (Figure 1.4a) supported a plesiadapiform–euprimate relationship to the exclusion of dermopterans. Dermoptera grouped with Chiroptera when bats were included and with Scandentia when Chiroptera was excluded from the analysis. The Dermoptera–Scandentia grouping is also found in many other recent molecular and morphological studies (Liu et al. 2001; Murphy et al. 2001a, b; Springer et al. 2003, 2004; Sargis 2002d, in press).

We have since expanded Silcox’s data matrix by including data from new plesiadapiform skeletons, new postcranial characters, and novel postcranial data from *P. lowii* (Bloch and Boyer 2003; see also Bloch et al. 2002, submitted; Figure 1.4b). This analysis supports a more basal position for micromomyids and a sister group relationship between Plesiadapoidea (including Plesiadapidae, Carpolestidae, Saxonellidae, and the basal form *Chronolestes simul* from Asia) and Euprimates. Thus, this analysis also supports the idea that plesiadapiforms were a paraphyletic stem group at the base of the order Primates. Plesiadapiforms,
Skeletons representing three plesiadapiform families were recovered from Late Paleocene limestones (Bloch and Boyer in press: fig. 3). Paromomyidae is represented by (a) *Acidomys hebeticus* (UM 108207) and (b) *Ignacius* cf. *I. graybullianus* (UM 108210). Carpolestidae is represented by (c) *Carpolestes simpsoni* (UM 101963; Bloch and Boyer 2002: fig. 2a). Plesiadapidae is represented by (d) *Plesiadapis cookei* (UM 87990). Scales = 5 cm
and specifically plesiadapoids, are crucial for establishing primitive states for Euprimates in this hypothesis.

One implication of these results is that plesiadapiforms might be best included in the order Primates (Silcox 2001, in press), in a return to earlier...
conceptions of how to define the group (Szalay 1975; Szalay and Delson 1979; Szalay et al. 1987). This notion has received some recent support from a previous proponent of the plesiadapiform–dermopteran clade (Kay 2003). The ecological scenarios discussed above take as their starting point the common features of modern primates. Plesiadapiforms do not possess all of these traits. With the recognition that plesiadapiforms constitute the primate stem lineage, discussing “primate origins” then involves dealing with at least two sets of evolutionary transitions—first, the branching off of the primate stem and evolution of the earliest primate (Purgatorius; Van Valen and Sloan 1965; Johnston and Fox 1984) and second, the origin of Euprimates. Earlier discussions of “primate origins” that explicitly endeavored to explain only the latter transition (Cartmill 1972; Rasmussen 1990; Sussman 1991) are inherently flawed in trying to account for the concerted evolution of character complexes that did not arise at the same time, mixing the effect of multiple evolutionary transitions. This is true even if one chooses to classify plesiadapiforms as a non-primate sister group to the order.

It is possible to formulate and test predictions about the sequence in which anatomical transformations occurred pursuant to the various ecological scenarios discussed above (Bloch et al., submitted) by the use of the hypothesis of primate supraordinal relationships given in Figure 1.4b. In the same way that “brains first” scenarios of human origins relied on the evolution of large brains before features for bipedalism, so the ecological scenarios of primate origins require a certain order for the addition of traits through time for them to be considered valid.

1.2.2.2 Predictions for ecological scenarios of primate and euprimate origins

Under the arboreal hypothesis, the prediction is inherent in the model that the evolution of characteristically primate traits coincided with a move into an arboreal habitus. If, on the other hand, the ancestors of Primates were already arboreal while lacking such traits, then the arboreal hypothesis would be effectively falsified. It would also be falsified if the evolution of characteristic primate features pre-dated a move to the trees, for example if forward facing orbits were found in an animal otherwise adapted for a terrestrial habitus.

The grasp-leaping hypothesis posits a relationship between the evolution of features for grasping with those for leaping. As such, if grasp-leaping is to function as an explanatory hypothesis for euprimate origins, then the evolution of these features should coincide in time. Visual features for improved stereoscopy should also coincide with the adoption of a more rapid, leaping,
locomotor mode. Although it may still be true that early euprimates were functionally grasp-leapers, if such a coincident evolution of the relevant traits is not found then this hypothesis would lose its explanatory power as a central motivating force in euprimate origins.

Cartmill (1992), in discussing the development of his visual predation hypothesis, wrote,

- Noting that marked optic convergence is also a characteristic of cats and many other predators that rely on vision in tracking and nabbing prey, Cartmill sought the adaptive significance of this trait in the predatory habits of small prosimian primates like Microcebus, Loris, and Tarsius, which track insect prey by sight and seize them in their hands. Grasping extremities and claw loss, he suggested, had also originated as predatory adaptations . . . (Cartmill 1992 p 107; emphasis ours)

This statement implies that these features should be tied together temporally, since they were acquired as part of the same adaptive shift to a more predatory pattern of behavior. Kirk, Cartmill, Kay, and Lemelin (2003 p 741b) claimed that Bloch and Boyer (2002) mischaracterized visual predation suggesting that “As originally formulated (Cartmill 1972), Cartmill’s thesis interprets the prehensile, clawless extremities of primates as adaptations for locomotion on slender arbo-real supports.” Regardless of these authors’ current opinions, this is directly contradicted by Cartmill’s own account (1992), quoted above, of what he meant when he formulated this hypothesis.

Since visual predation involves an increasing reliance on insect prey, this should also be reflected in the teeth of the earliest euprimates. This is true even if these forms were grasping prey with their hands rather than teeth (contra Cartmill 1972, 1974) because dental features for insectivory reflect not only prey capture but also processing of food items with the unique physical properties of insects. If, however, early primates or euprimates were found to be equally or less insectivorous than their forbears, or a disassociation were to be found between grasping and visual traits, then visual predation would be refuted as a central motivating force in early primate evolution. Since visual predation also relies to some degree on nocturnality (Allman 1977; Cartmill 1992), a finding that the earliest primates or euprimates were diurnal would substantially weaken this hypothesis.

The angiosperm diversification hypothesis predicts two stages in the evolution of primates. First, with the initial exploitation of the arboreal mixed feeding adaptive zone, a dental shift reflecting more use of plant resources should be seen. Second, as the terminal branches were invaded and the use of the food resources from this milieu was intensified, grasping and dental features reflecting these changes should appear. Disassociation between dental traits for eating fruit or
flowers, and those indicating the ability to access terminal branches, would weaken the explanatory power of this hypothesis.

Similarly, the combined hypothesis involving first terminal branch feeding on fruit, and then visual predation, suggested by Rasmussen (1990), requires that “the earliest euprimates had grasping feet and blunt teeth adapted for eating fruit, but retained small, divergent orbits like those of *Plesiadapis*” (Cartmill 1992 p 111). Subsequent evolution should add features for visual predation, such as forward facing orbits and teeth with improved capabilities for processing insects, to this basic model. If, however, convergent orbits evolved at the same time as grasping feet or blunt teeth, or their appearance was not coincident with the evolution of teeth better designed for eating insects, then this model would be effectively falsified.

### 1.2.2.3 Assessment of ecological scenarios

With the well-supported pattern of relationships found by the current authors (Figure 1.4b), it becomes possible to consider the predictions outlined above in light of what we know about the fossil record (Figure 1.5). In terms of the arboreal hypothesis, the inferred arboreal habits of all plesiadapiforms known from postcranials (Szalay and Decker 1974; Szalay et al. 1975; Szalay and Dagosto 1980; Szalay 1981; Szalay and Dradhorn 1980; Beard 1989; Gingerich and Gunnell 1992; Boyer et al. 2001; Bloch and Boyer 2002, 2003, in press) make it clear that the ancestors of Euprimates were already arboreal. This is further indicated by the inclusion of Primates in Archonta or Euarchonta because this supraordinal group likely had an arboreal ancestor (Szalay and Dradhorn 1980; Sargis 2001a, 2002e). As such, distinctively euprimate traits cannot be linked to a simple move from a terrestrial to an arboreal habitus. The fact that *P. lowii*, an arboreal mammal, may be the best living model for the ancestor of Archonta or Euarchonta and possibly for Primates (Sargis 2001a, 2002e; Bloch et al. 2003), strongly suggests that arboreality is a feature that evolved prior to the base of the primate radiation (Szalay and Dradhorn 1980). Many of the features that have been cited as possible archontan or euarchontan synapomorphies (Szalay and Dradhorn 1980; Szalay and Lucas 1996; Sargis 2002d; Silcox et al. 2005) can also be linked to arboreal locomotion. It is likely that arboreality evolved in the ancestor of Euarchonta (Szalay and Dradhorn 1980; Sargis 2001a, 2002e), and that this trait was retained (but did not originate) in the ancestor of Primates.

The evolution of grasping is central to the assessment of both the graspleaping and visual predation hypotheses. With a better fossil record for plesiadia-piforms, it is now clear that grasping is not a single character state or set of
coordinated transformations. The ancestral euarchontan was likely capable of *Ptilocercus*-like grasping (Szalay and Dagosto 1988; Sargis 2001b, 2002b, e, 2004). As Bloch and Boyer (2002, 2003) demonstrated, the evolution of fully euprimate-like grasping was at least a two-stage process. Features for manual grasping, including relatively long digits of the hand, are present in all plesiadapiforms known from relevant material, with the exception of plesiadapids who have secondarily lost this trait (Bloch and Boyer 2002, 2003; Boyer et al. 2004). Euprimate-like pedal grasping, including a divergent big toe with a nail, is present in *Carpolestes simpsoni*, and can be reconstructed as having evolved in the common ancestor of Plesiadapoidea and Euprimates (Bloch and Boyer 2003).

No plesiadapiforms known show any features associated with specialized leaping (Szalay et al. 1975; Szalay and Dagosto 1980; Beard 1989; Gingerich and Gunnell 1992; Bloch and Boyer 2002, in press). *Carpolestes simpsoni*, for example,
lacks the relatively long legs typically seen in a leaping mammal, indicating that it was a more generalized arboreal quadruped (Figure 1.3; Bloch and Boyer 2002). The first primate taxa with clear leaping characteristics are early euprimates such as Cantius and Omomyx (Rose and Walker 1985; Anemone and Covert 2000). Although current evidence suggests that leaping and visual traits did evolve at the same time, there is a distinct offset between the evolution of features related to grasping and those for leaping. In light of this, although early euprimates were likely grasp-leapers and leaping may have evolved in the ancestral euprimate (Szalay and Dagosto 1980, 1988; Dagosto 1988), the evolution of grasp-leaping was not the event that shaped the origin of primates or euprimates.

Similarly, no known plesiadapiforms show any of the specialized features of the orbital system that are associated with euprimate-like vision, including a complete postorbital bar, convergent orbits, reduced snout, or an enlarged and reorganized brain (Russell 1964; McKenna 1966; Szalay 1969, 1972; Kay and Cartmill 1977; Kay et al. 1992; Bloch and Silcox 2003, 2006). Therefore, there is an offset between the evolution of grasping and visual features. As such, their coordinated acquisition as part of a shift to a new mode of feeding, visual predation, was not the decisive event in shaping primate or euprimate origins.

It is still possible that adding visual predation to the behavioral repertoire of euprimates was an important event in the evolution of this group, in which grasping features effectively acted as an exaptation (=preadaptation). However, there is evidence that leads one to doubt this scenario, however. First, as discussed above, visual predation becomes mechanistically implausible if the earliest euprimates were diurnal (Allman 1977; Cartmill 1992). Ni et al. (2004) recently published a primitive euprimate skull from Asia that they interpreted as having been diurnal. Although there are some problems with this conclusion (Heesy and Ross 2004; Martin 2004; Bloch and Silcox 2006), it draws attention to the fact that the ancestral activity period for euprimates remains a matter of debate. Second, if euprimates did undergo a transition to becoming more focused on visual predation, then they should have teeth that are indicative of a more insectivorous diet than their precursors. This is not demonstrably true. The earliest known euprimate, Altiatlasius koulchii, has extremely low-crowned teeth with very bunodont cusps (Sige et al. 1990), which is not consistent with a predominantly insectivorous diet. Of the two best-documented groups of early euprimates, adapids are usually viewed as being frugivorous and omomyids as omnivorous or frugivorous (Rose 1995; Strait 2001). The only gut contents known for a primitive fossil primate (from the adapid Godinotia from Messel) include fruit remains and no insects. This is likely a real reflection of diet rather than a taphonomic artifact because in other Messel specimens insect remains preserve well (Franzen and Wilde 2003).
Cartmill (1972) attempted to counter this objection by arguing that features such as lower crowned teeth and a reduced stylar shelf, interpreted as being associated with a more herbivorous diet (Szalay 1968), do not “... contradict the hypothesis that the ancestral primates were primarily insectivorous” (Cartmill 1972 p 117). Although it is certainly very probable that early euprimates ate some insects, as do all modern small primates (Fleagle 1999), this observation is irrelevant for two reasons. First, the key information for assessing evidence for an evolutionary transition is not static reconstruction of diet but verification of a change in diet. Just demonstrating that early euprimates ate some insects is not sufficient—if visual predation is to be supported as an important formative force in euprimate evolution, a shift to greater insectivory must be demonstrated. Unambiguous evidence for such a shift is lacking from the fossil record.

Second, the relevant comparative sample for early euprimates is not living primates but contemporary insectivores. If early euprimates were succeeding and diversifying primarily because they were improving their insect-harvesting abilities, then they should show dental features that indicate that they were at least as well adapted for processing insects as insectivorous mammals living at the same time. This is not the case—most specialized insectivores from the Paleocene and Eocene have much higher crowned teeth and sharper cusps than early primates. It is precisely the absence of such features and the presence of characteristics for processing non-leafy plant material, such as low-crowned molars with broad talonid basins, which make it possible to separate primitive primate and insectivoran teeth in the fossil record. The insectivorans most similar to primates in dental form can be reconstructed as having a more omnivorous diet than their specialized insectivorous kin. For example, the erinaceomorph *Macrocranion tupaiodon* from Eocene deposits at Messel, which has superficially primate-like teeth, is known from stomach contents to have eaten not only insects but also plant material and substantial quantities of fish (Storch and Richter 1994).

The features suggesting a more herbivorous diet in early primates and euprimates are supportive of the angiosperm diversification hypothesis. Since fully euprimate-like pedal grasping pre-dates the origin of Euprimates, the pattern of acquisition of adaptive features does not match precisely the timing of events envisioned by Sussman (1991; Sussman and Raven 1978). Nonetheless, his main points do still have merit. The origin of the order Primates is associated with dental changes for increased herbivory (Szalay 1968). This was likely associated with specializations for terminal branch feeding, including grasping traits, in an omnivorous animal. Within various plesiadapiform lineages and early euprimate groups, improved features for exploiting plant propagative organs continue to appear through the Paleocene and Eocene (Gingerich 1976; Biknevicius 1986; Rose 1995; Bloch and Boyer 2002).
The one major shortcoming of the angiosperm diversification hypothesis is that it may fail to explain the rest of the distinctive traits seen in modern primates (i.e., visual features and leaping). Sussman (1991 p 219) did try to build an explanation for visual traits into his ideas by linking the manipulation of small food items in low light conditions and on tenuous substrates to a need for “acute powers of discrimination and precise co-ordination.” The lack of evidence in modern analogues such as bats for a solid link between frugivory and convergent orbits has led to this view largely being discounted (Ravosa and Savakova 2004). However, perhaps modern models, such as bats, are not appropriate for comparisons to early euprimates, who undoubtedly used a different feeding mode. As Crompton (1995 p 18, emphasis his) pointed out, euprimate visual traits may have evolved for their power in breaking crypsis (i.e., picking out camouflaged food items from the background), and as such could have been beneficial in allowing detection of the “small, and often very inconspicuous” food items taken by small primates. In terms of leaping, Rasmussen’s (1990 p 273) observations of Caluromys also offer a potential explanation for the value of this locomotor mode to a terminal branch feeder: “The grasping and leaping acrobatics exhibited by C. derbianus in Costa Rica enabled them to gain access to fruit that was apparently off limits to most of the other nocturnal frugivores of the study area.” Perhaps it was refinements to terminal branch feeding techniques, offering new access to previously inaccessible food sources and greater abilities for discriminating food choice, which marked the transition to Euprimates. This would only be a difference from Sussman’s scenario in the traits that are emphasized, not in the ecological events that were key.

Rasmussen’s combination hypothesis could be seen as offering an alternative to the angiosperm origins scenario that explains first the grasping and fruit-eating dental features of basal primates and then the visual characteristics of euprimates. However, this combination hypothesis suffers from the same problems as visual predation in linking the evolution of orbital traits to increased insectivory in the absence of evidence for such a dietary shift. On the basis of the current evidence, the angiosperm diversification hypothesis applies best to the evolution of early primates. Furthermore, Szalay’s (1968) view of the key event in primate origins being a dietary transition to a more plant-dominated repertoire is also supported by the current evidence.

1.3 Timing and place of origin of primates and euprimates

The earliest occurring primate known is Purgatorius, from either the latest Cretaceous (Van Valen and Sloan 1965; Van Valen 1994) or earliest Paleocene
Primate origins and supraordinal relationships: morphological evidence

(Johnston and Fox 1984; Lofgren 1995) of North America. Most of the rest of the early primate fossil record is North American, including all definitive micromomyids and palaechthonids, most microsyopids, and all the most primitive paromomyids, carpolestids, plesiadapids, and possibly saxonellids (Fox 1991). Plesiadapiforms have only been known from Asia since 1995 (Beard and Wang 1995), which suggests that this geographic bias may be a sampling phenomenon. However, the lack of Asian (or European or African) taxa as primitive as *Purgatorius* is supportive of a North American origin for the order Primates (Bloch et al. submitted).

In terms of the place of origin of Euprimates, Beard (1998) argued that it could be reconstructed as unequivocally Asian. We view this assertion as excessively confident. Silcox (2001) reached a very different conclusion, with origins in Asia, Africa, North America, or even Europe being possible in the context of the current record. Causes for this equivocation include the African location of the earliest known euprimate, *Altiatlasius koulchii*, the Asian location of the primitive euprimate *Altanius orlovi*, the North American location of much of the primitive plesiadapoid and euprimate record, and the European location of both some early euprimates (*Donrussellia*) and of the poorly sampled plesiadapiform family Toliapinidae, which may be related to early euprimates (Silcox 2001; Figure 1.4a).

The time of origin of the two major clades under discussion here (Primates and Euprimates) can only be minimally constrained using fossil data. As noted above, the earliest known primate, *Purgatorius*, is approximately 65 Myr old. In light of the primitive nature of this taxon, the fossil record is not consistent with a date much earlier than this, putting the origin of the group in the latest Cretaceous, somewhat later than even the most conservative molecular estimates (Springer et al. 2003). The earliest occurring euprimate, *Altiatlasius koulchii*, is Late Paleocene in age (Sige et al. 1990; Gheerbrant et al. 1998), implying a divergence for euprimates before the Early Eocene. Furthermore, since the sister group to Euprimates (i.e., Plesiadapoidea) had diverged from their common stem by the latest Early Paleocene, Euprimates must be at least that old.

1.4 Conclusions: what is a primate? (Coda)

When Cartmill developed the visual predation hypothesis, he suggested the removal from Primates of any taxa that lacked modern primate-like orbital and grasping features, and thus presumably had not used this mode of feeding (Cartmill 1972, 1974, 1992). This was the primary basis for Cartmill’s suggested removal of plesiadapiforms from Primates. Such an approach to defining Primates was perhaps an overoptimistic view of the support for visual predation—if,
as suggested here, an evolutionary transition to this pattern of behavior is not clearly indicated by the fossil record, then this is surely not an appropriate criterion by which to determine inclusion or exclusion of taxa in the order Primates. This view is underscored by the fact that of the three “ordinally diagnostic” traits that Cartmill (1972 p 121) named to diagnose a plesiadapiform-free order Primates—“... the petrosal bulla, complete postorbital bar, and divergent hallux or pollex bearing a flattened nail ...”—two are now known in plesiadapoid plesiadapiforms (Bloch and Boyer 2002, 2003; Bloch and Silcox 2006). The fossil record demonstrates that the characteristic primate traits listed in the introduction arose in a step-like fashion (Figure 1.5). Thus, the criterion that all of these features must be present in a particular taxon for it to be considered a primate is biologically unnatural. Doing so would exclude taxa on the primate stem who have some, but not all, of these traits, but who postdate the divergence of the primate lineage from the rest of Mammalia. As demonstrated above, such stem taxa are critical for understanding the origin and early evolution of Primates, as well as the accumulation and modification of crucial features within this lineage.

As advocates of phylogenetic taxonomy have made clear, there are some distinct advantages to formal taxonomic definitions that are based on specifying a particular ancestor rather than on a list of mutable characters (Rowe 1987; De Queiroz and Gauthier 1990; Silcox in press). For this reason, although compiling lists of distinctive primate traits is useful to the process of understanding primate origins, it is inappropriate to consider them formal definitions. Using the precepts of phylogenetic taxonomy, Silcox (in press) suggested the following definition for Primates: “the clade stemming from the most recent common ancestor of Purgatorius and Euprimates.”

New discoveries will almost certainly change our views on the early parts of primate evolution. There are a number of substantial holes in the fossil record for primate origins, which when filled may fundamentally shift our perceptions of primate evolutionary history. First, there is a sizeable spatial discontinuity in the fossils currently available. Plesiadapiforms have only been discovered in Asia in the last 10 years (Beard and Wang 1995). Since we consider Altiatlasius to be a euprimate, the only potential plesiadapiforms from Africa are the poorly known Azibius and Dralestes (Tabuce et al. 2004). For early euprimates, the few specimens of primitive forms known from Asia and Africa are suggestive of a much larger radiation that is almost completely unknown (Silcox 2001). Even in North America, the geotemporal patterning of the plesiadapiform and euprimate fossil records means that there are still substantial areas at crucial times that remain unsampled.
Second, some taxonomic groups are also undersampled. Two families of plesiadapiforms, Palaechthonidae and Toliapinidae, have the potential to be crucial to an understanding of early primate and euprimate evolution, but both are very poorly known. The best-known plesiadapoids are all relatively derived members of their respective families. In light of the important position of Plesiadapoidea, as the sister taxon to Euprimates, finding more, and more complete, primitive plesiadapoid specimens is vital (Boyer et al. 2004). Perhaps most importantly, a gap still exists between the known plesiadapiforms and the earliest euprimates. No known plesiadapoid has the morphology that would be expected in a euprimate ancestor—they are all too derived in features such as dental reduction, enlargement of the anterior-most incisors, and/or the shape of P4. Because the earliest plesiadapoids are late Early Paleocene in age, Euprimates must have a ghost lineage stretching through the Middle and Late Paleocene, which is entirely unsampled. Filling this particular gap will be central to clarifying the evolutionary and adaptive significance of traits for euprimate-like vision and leaping. In light of the complete absence of taxa to fill this gap from the comparatively well-sampled North American record, it seems most plausible that they were living in the Old World.

Finally, since understanding the supraordinal relationships of Primates is central to reconstructing events at the base of the order, a better fossil record for other euarchontan groups is also central to the problem of primate origins. As it stands, the Paleogene fossil record for scandentians and dermopterans is virtually nonexistent, with the exception of a single dermopteran specimen from the Eocene of Thailand (Ducrocq et al. 1992), fragments of scandentian teeth from the Eocene of China (Tong 1988), and plagiomenids, which may be fossil dermopterans. Furthermore, a better understanding of various other fossil groups for whom a tie to Archonta, or specifically to Primates, has been suggested (apatemyids, nycitheriids, mixodectids; Szalay and Lucas 1996; Hooker 2001; Silcox 2001; Silcox et al. 2005), has the potential to further clarify the evolutionary events downstream from Primates in the euarchontan evolutionary tree.

Although this discussion of holes in the fossil record may seem disheartening, the enormous progress that has been made in the last 10 years for our understanding of primate origins suggests we may not have to wait long for some of these holes to be filled. We have moved from a position analogous to that of early anthropologists arguing about whether brains or bipedalism arose first in human evolution, without having any relevant data to choose between the two, to being able to actually test hypotheses about the order of acquisition of traits in early primate evolution. We can only hope that continuing diligence on the part of researchers interested in primate origins will serve to fill some of these
gaps and allow us to continue expanding our knowledge of the earliest chapters in our own evolution.

Acknowledgments

Our thanks to P.D. Gingerich, K.D. Rose, F.S. Szalay, A.C. Walker, and J.G. Fleagle for conversations relevant to this chapter. We thank Annette Zitzmann for providing the photo of *Ptilocercus* in Figure 1.1. Research was funded by grants from NSERC, Wenner-Gren, the Paleobiological Fund, Sigma Xi, NSF (SBR-9815884), and the University of Winnipeg to MTS; NSF (BCS-0129601) to G.F. Gunnell, P.D. Gingerich, and JIB; NSF (SBR-9616194), Field Museum of Natural History, Sigma Xi, and the Yale University Social Science Faculty Research Fund to EJS; 2002 NSFGRF to DMB.

References


Bloch JI, Silcox MT, Boyer DM, Sargis EJ (submitted) New Palaeocene skeletons root the primate tree
Nouvelles données sur les mammifères du Thanetien et de l’Ypresien du Bassin d’Ouarzazate (Maroc) et leur contexte stratigraphique. Palaeovertebrata 27: 155–202


Le Gros Clark WE (1959) The antecedents of Man. Quadrangle Books, Chicago


Primate origins and supraordinal relationships: morphological evidence


Sargis EJ (2002c) A multivariate analysis of the postcranium of tree shrews (Scandentia, Tupaiidae) and its taxonomic implications. Mammalia 66: 579–598
2 Molecular Evidence on Primate Origins and Evolution

Hans Zischler

Abstract

Over the last decade, research on primate phylogeny has been of increasing interest to the scientific community. From the perspective of molecular evolution, this is mainly due to the fact that the mass generation of molecular sequences has become easy and cost effective. With the generation of complete sequences for several eutherian organisms including humans and the mouse, a well-accepted phylogenetic interpretation for all members of the Euarchontoglires and all major groups of the primate order is feasible and would represent a new starting point for meaningful comparative research. Such a phylogenetic framework would link humans with the mouse, which is generally regarded as the main eutherian model organism. Thus, our knowledge of primate origins and the evolution of primates is a prerequisite for a postgenomic era in which aspects of functional genetics and character evolution will form a focal point of genetic research. Despite the pace at which primate sequences can be generated in whole genome shotgun (WGS)-sequencing projects, primate origins as well as several branching events in primate divergence remain far from settled. First, complete primate genome sequences are currently available for two representatives of the Old World monkeys and hominoids and humans only. Information is lacking on the deeper primate splits and comparative data are restricted to parts of primate genomes (ENCODE project). Second, it is obvious that the peculiar mode of sequence evolution (including gene-, lineage-, and position-specific evolutionary rates), combined with deep splitting events that often occurred during small time intervals may possibly lead to incongruence between gene and species trees. To avoid this, it will be necessary to have enormous amounts of sequence data and the implementation of more realistic assumptions about sequence evolution models in sequence-based phylogenetic tree reconstructions. Moreover, alternative molecular approaches, including both the incorporation of data of so-called “rare genomic changes” (RGCs) and a combination of both neontological and paleontological morphological data in total evidence approaches, are likely to contribute considerably to a firm
interpretation on the origin and evolution of primates. Below I summarize and discuss molecular evidence obtained for the origin and evolution of pri-
mates, stressing the potential of the inclusion of “RGCs,” mainly retropositions of short interspersed nuclear elements (SINEs) in this context.

2.1 Introduction

2.1.1 Primate genetics in the postgenomic era

As the closest extant relatives of *Homo sapiens*, nonhuman primates have always been of particular interest in biological and biomedical research. With the advent of the first drafts of the complete DNA sequence of the human genome reported in 2001 by the IHGSC and Celera Genomics, the postgenomic era began, focusing on functional and diversity aspects of human DNA. With the complete genome sequence of the mouse, generally considered the most important eutherian model organism, the goal of pinpointing the evolutionary changes in the lineage to humans and of bridging the gap between eutherian model systems and humans has brought the order Primates even more into the focus of postgenomic evolutionary research. Obviously, the phylogenetic history of nonhuman primates forms a part of our own evolutionary history. Though each primate taxon represents the realization of a stochastic evolutionary process that occurred during primate divergence, tracing the genomic changes that can be identified by comparing different eutherian and primate taxa will shed light on the processes that can shape our genomes over evolutionary time scales. As with every analysis, e.g., of morphological, physiological, or behavioral character evolution, the latter requires an undisputed phylogenetic framework for discerning between homology and analogy. For that reason, this review will mainly summarize the molecular data and interpretations currently to hand on the origins and phylogenetic affiliations of primates and closely related eutherian orders. Although a complete appraisal of the respective available information would clearly go beyond the scope of this review, molecular evidence on the divergence of the main primate groups will be discussed. Though a major research focus that is attracting enormous public interest is the understanding of human uniqueness, Goodman et al. (2005) correctly stated that this understanding cannot be complete if we look exclusively at the (molecular) differences between humans and their phylogenetically closest primate relatives. To fully comprehend human uniqueness, we need to understand how humans differ from chimpanzees, how great apes differ from the remaining haplorhines, how strepsirhines are set apart from the latter, and how primates differ from other eutherians.
A special emphasis will be given to a type of marker system that allows a molecular cladistic approach to determine sister group relationships by analyzing rare genomic changes (RGC) such as insertions of retroposons with respect to their presence or absence in certain taxa. Though focusing on the molecular cladistic information of these sequence changes, it has to be kept in mind that they might exert an effect on the respective transcriptomes and proteomes.

Although mainly molecular evidence will be presented in this review, it must be kept in mind that paleontological and neontological morphological data are urgently required to complement the molecular database and to establish a “total evidence” phylogenetic framework. This will in turn enable us to trace character evolution during the emergence and divergence of primates. Remarkably, the overall congruence of phylogenetic interpretations based on the two sources of data is substantial, offering the prospect that total evidence approaches will soon lead to a broadly accepted hypothesis on the eutherian phylogenetic affiliation, origin, and evolution of primates.

2.1.2 Eutherian evolution: molecular data and problems

Until the advent of cost-effective generation of molecular data, research on mammalian phylogeny was essentially dominated by morphological and paleontological approaches.

Recently, Novacek summarized a consensus tree from these analyses that had its roots in the influential work of Gregory and Simpson dating back to the beginning and middle of the last century. It has, however, to be stressed that many of the phylogenetic constellations proposed therein were and are still matters of debate among paleontologists and morphologists. On the basis of the classical approach, Xenarthra (e.g., armadillos, anteaters) constitutes the most basal placental group, with the remaining eutherian orders comprising the Ungulata, Anagalida, and Archonta. The latter group or superorder comprises the orders Scandentia, Dermoptera, Chiroptera, and Primates. Over a decade ago, sequences for complete mitochondrial genomes became available, adding molecular evidence to a discussion that formerly had exclusively dealt with interpretations based on nonmolecular data.

After the draft human genome sequence was completed in 2001, plans to expand whole genome shotgun (WGS) sequences to other mammalian species were proposed and projects initiated. Besides the complete human, chimpanzee, and rhesus monkey sequences, other non-primate and primate representatives including the bush baby, the common marmoset, and the tree shrew are about to be sequenced completely (a constantly updated list of the sequencing proposals
and projects can be viewed under http://www.genome.gov/10002154). In addition, sequencing projects, especially the encyclopedia of DNA-elements-project (ENCODE project, Encode Project Consortium 2004), aim to create a large set of comparative data mainly from functional regions of the genome. Numerous taxa are in the pipeline for partial sequencing to be carried out from bacterial artificial chromosome (BAC) clones already available for a rather representative set of primate taxa (see http://bacpac.chori.org/). A large-scale comparative sequencing of the cystic fibrosis transmembrane conductance regulator (CFTR) locus region can be regarded as an initial example of such an undertaking (Thomas et al. 2003). Thus any deficiencies in the molecular data, mostly involving taxonomic representation and suitability for eutherian phylogenetic analyses, are likely to be erased in the foreseeable future. Initially, phylogenetic interpretations of complete mitochondrial DNA (mtDNA) from different eutherian taxa led to various unorthodox phylogenetic proposals, provoking a debate about the appropriateness of mtDNA-based eutherian phylogenetic tree reconstructions or even molecular approaches in general. However, over the last years a molecular consensus has been emerging, with a topology that deviates in essential aspects from the morphological tree, due to an improvement of bioinformatic tools that incorporate more detailed knowledge of sequence evolution models. Second, the flood of incoming sequence data—mainly from the nuclear side—is overcoming the major early problems of molecular studies, namely an incomplete and unbalanced taxon sampling combined with limited sequence information from short DNA alignments. Recent compilations of extended molecular data sets suggest three novel superordinal clades, and that extant eutherians can be partitioned into four major groups, the Laurasiatheria, Xenarthra, Afrotheria, and Euarchontoglires. This grouping emerged at a 1999 meeting (Waddell et al. 1999) and was corroborated in several subsequent studies (Madsen et al. 2001; Murphy et al. 2001a, b). The Euarchontoglires comprise lagomorphs and rodents on the one side and the scandentians, dermopterans, and primates on the other. Interestingly, both mtDNA and nuclear DNA (nDNA) analyses reproducibly exclude the chiropters from the Euarchontoglires, suggesting a closer affiliation of chiropters to, e.g., cetartiodactyls than to primates. As to phylogenetic affiliations inside the Euarchontoglires, the rodents and lagomorphs are considered to represent sister groups. But the phylogenetic affiliations among the remaining members, the primates, flying lemurs (Dermoptera), and tree shrews (Scandentia) remain highly disputed. Solid support for these superordinal groups has come from independent studies of the substitutional evolution of nuclear genes and complete mtDNAs. Many of these studies may be hampered by reconstruction problems due to position, gene, lineage-specific evolutionary rates, uneven nucleotide compositions, and inadequate taxon sampling. In fact, powerful
molecular cladistic information can be found in RGCs that convincingly corroborate the four eutherian clades (Springer et al. 2004; Kriegs et al. 2006). Mainly because point mutations can be regarded as weak, homoplasy-prone characters, the recognition of multiple phylogenetic characters including RGC has gained increasing popularity in phylogenetic research.

Under the umbrella of RGC, an assemblage of completely different molecular scenarios is found. These include indels in protein coding regions—an RGC that supports the Euarchontoglires-clade—as well as changes in gene order and intron position, large-scale rearrangements and duplications, and finally the existence and distribution of transposable elements in different genomes.

It has to be stressed that our current knowledge on the different DNA turnover mechanisms that stand behind these phenomena is very variable. Hitherto, probably the best understood RGC is represented by the transpositional activity of repetitive DNA. Due to the abundance of these potential marker systems in typical metazoan genomes, it is currently the most widely applied RGC and will be presented in more detail below.

2.1.3 Problems relating to primate origins and phylogenetic affiliation to other eutherians

From a paleontological point of view, the definition of primate origins and separation of proto-primates from closely related eutherians is not a trivial task. First, there are no very specific morphological fossil features that can unify all primates. Not taking molecular data into account, primates are characterized by a long list of common morphological features given by Le Gros Clark in 1959 and supplemented by Napier and Napier (1967). Superficially, this list of primate features should enable researchers to clearly define an individual that belongs to the primate order. However, there is no unique morphological, physiological, or behavioral character that clearly characterizes a primate. The criteria mentioned in the references above rather represent a collection of symplesiomorphic features and do not contain an undisputed synapomorphy defining a clade. Many “typical primate” features are either behavioral or depend on soft tissue anatomy, so deciding whether this or that fossil represents a primate is notoriously difficult.

Second, the scanty fossil record of early primates represents a major problem for studying the origin of primates. Tavaré et al. (2002) suggest that the currently available fossil record only shows 7% of all primate species that ever existed. They estimated that the common primate ancestor existed about 81.5 million years ago (Ma), which is considerably older than previous estimates based, e.g., on paleontological evidence. For the latter, Purgatorius—a taxon that lived in North
America and Eurasia at least 65 Ma during the Late Cretaceous and Early Paleocene—was hitherto regarded as the prime candidate on which all extant primate species coalesce (Shoshani et al. 1996).

Another complicating factor is represented by growing evidence that archaic primates or Plesiadapiformes are much more diverse than previously assumed. Various phylogenetic constellations have been proposed and are still being discussed for the different plesiadapiform families and extant members of the euarchontan orders. Molecular data obtained from extant taxa might thus complement this discussion by enabling hypotheses on the phylogenetic affiliations between primate and non-primate crown groups.

Before I review the available molecular information on primate origins, it is necessary to provide some information on retroposable DNA sequences. Evidence from this type of RGC is currently used to corroborate phylogenetic interpretations based on the analysis of substitutional sequence evolution. To estimate the probability of homoplasies and reversals and to assess the value of this molecular cladistic marker system, some basic knowledge of the biology of retroposing DNA is necessary.

### 2.2 Molecular biological approaches and methods

#### 2.2.1 Genetics of retroposons

As in other eutherian mammals, primate genomes harbor a variety of repetitive sequences with no obvious and global function (Smit and Riggs 1995). The bulk of these discernible sequences, constituting more than 40% of the human genome (Li et al. 2001), represents remnants of molecules that were mobilized during the evolution of extant species. This number is even likely to be an underestimate of the repetitive sequences in primate genomes, since sequence evolution might eradicate signatures of formerly repetitive DNA of extinct taxa. More WGS sequence data from primates are needed before we can conclusively answer the question of the quantity of each specific repetitive sequence in a certain taxon.

Sequences that belong to class II elements can be mobilized by a cut and paste mechanism involving a DNA intermediate, while class I elements transpose replicatively by a copy and paste mechanism via an RNA intermediate (retro-elements). Class I, or retroelements, decoding their own reverse transcriptase, allows an autonomous transposition. This group includes the non-LTR (long terminal repeat) elements such as long interspersed elements (LINEs) decoding two proteins, an RNA-binding protein and an endonuclease/reverse transcriptase, and LTR elements as well as the true retroviruses (Hedges and Batzer 2005).
Short interspersed elements (SINEs) form another type of class I element with nonautonomous transposition, exploiting instead the enzymatic retropositionary machinery—proteins with reverse transcriptase and endonuclease activity—of other elements such as LINEs. SINEs are the most abundant class I sequences to be traced in primate genomes. Typically, SINEs range between 75 and 500 bases in size and may be amplified to a copy number well beyond $10^4$ total copies per genome. The majority of the eukaryotic SINE families described so far can be traced back to a tRNA ancestor, with a minority apparently derived from 7SL RNA. A common feature of both types of SINEs is their internal promoter that is specific for RNA polymerase III (Okada 1991a, b). The most abundant SINE in primates is represented by the so-called Alu-SINE, which evolved from a 7SL RNA.

Further actively retroposing sequences are represented by the LINE retroposon family. Considering the length and copy number of LINEs, these sequences represent the most abundant sequences in the human genome. Primate L1 sequences amount to roughly 6,000 bp in length, containing an RNA Pol II promoter, two ORFs, a 3′ UTR, and Poly-A tail. An RNA-binding protein is decoded in the first ORF; the second ORF codes for an endonuclease and reverse transcriptase. Of more than 800,000 estimated L1 copies in the human genome, most of the LINEs in hitherto sequenced primate genomes are 5′ truncated (and therefore also lack the capacity of further retroposition). Thus close to 10,000 copies are full length and about 100 are active for retrotransposition in a tissue culture assay. It is estimated that over one-third of mammalian genomes are the result, directly or indirectly, of L1 retrotransposition (Han and Boeke 2005). Integration of non-LTR retrotransposons, such as mammalian L1 and related SINEs (e.g., Alu sequences), is mediated by the mobilization machinery from LINEs represented in the two LINE–ORFs decoding an RNA-binding protein and an endonuclease/reverse transcriptase, respectively.

During SINE/LINE retroposition, the RNA intermediate is reverse-transcribed by a process called target-primed reverse transcription (TPRT) (Luan et al. 1993; Feng et al. 1996; Jurka 1997; Kazazian and Moran 1998; Kajikawa and Okada 2002). Here the endonuclease cuts the target DNA at two nicking sites, creating a staggered end break of the double stranded target DNA. This staggered end break is subsequently filled and is recognizable as a direct repetitive sequence flanking the integrated element.

Bioinformatic analyses of these direct repeats, representing the target sites, suggest that from the two nicking sites of this machinery only the 5′ site is relatively conserved in contrast to the 3′ site. Furthermore, it is not understood which factors affect the cleavage at the second site and how specific this process is for different combinations of SINEs and LINEs. Taken together, this is indicative
of the uniqueness of a single LINE-mediated retroposition even over evolutionary
time scales (Zingler et al. 2005).

As is the case for Alu-SINEs, also L1 sequences are obviously to a greater
extent mobilized in the germline, with a possible bias toward the male germline.
This phenomenon is currently seen as a consequence of the host defense toward
transpositional activity, which is mediated by the methylation of DNA. In the
course of germ cell development demethylation takes place, allowing increased
transpositional activity (Hedges and Batzer 2005).

Besides Alu sequences and L1 that altogether constitute more than 30%
of the human genome, other TE are still actively retroposing in primate
genomes and can be traced in high copy number in primate genomes. Prominent
SINE examples are the tRNA-derived galago monomer (Daniels and Deininger
1991) found in all lorisiform species analyzed to date (Roos et al. 2004) and
type II SINEs, the latter representing a SINE composed of both tRNA- and 7SL-
derivatives.

Further primate TE types include DNA transposons, SINE-R, LTR retro-
transposons, and endogenous retroviruses (ERVs). Apparently, a DNA transpo-
son TIGGER gave rise to smaller miniature inverted repeat transposable elements
(MITEs) sequences in an ancestral primate. Their exact quantitative contribution
to extant primate genomes remains to be determined. However among the
Euarchontoglires, DNA transposons that mobilize by cut and paste mechanisms
are apparently more commonly found in rodent genomes (Hedges and
Batzer 2005).

ERVs or remnants of them that still reside in primate genomes represent the
outcome of germline retroviral infections that took place at various time points
during primate divergence. In humans these ERV sequences constitute about 1%
of the genome. There are several lines of evidence that ERVs can alter the
expression pattern; their impact on the primate genomes was recently summa-
ized by Sverdlov (2000).

Finally, a composed element—the SVA repeat comprising Alu-like, SINE-R
and VNTR elements—is sometimes mentioned as a still active retroelement. As
among Alu-SINEs, the SVA is dependent on L1 mobilization factors. Both
chimpanzee and human genome analyses reveal that this element is still active
and also disease related. Obviously, not all types of repetitive DNA elements are
still mobilizable. The most prominent example in primate genomes is represented
by the mammalian wide interspersed repeats (MIRs). These MIR sequences are
highly divergent, suggesting a deep coalescence. Indeed, the vast majority of
active MIR sequences was actively transposing during early mammalian evolu-
tion, representing a shared ancestral character state among all eutherians (Hedges
and Batzer 2005).
2.2.2 SINEs as cladistic markers

Much research has been directed at the biology, genetics, and the genome-shaping as well as functional implications of retroposable elements. Retrotransposition-mediated deletions, gene conversion, or SINE-mediated recombination events that trigger further large-scale genomic rearrangements are known as directly shaping primate genomes. Furthermore, SINEs have repeatedly been shown to display a gene regulatory function from the epigenetic level to their role as possible modifiers of splicing events. Mobile elements are capable of moving around regulatory sequences for different transcription machineries as well as unrelated sequences 3′ to the repeat element—a process called transduction. Repeats can be exapted as functional parts of newly invented proteins thus exerting an effect on the proteome.

As temporal landmarks of evolution, SINEs and LINEs have also gained widespread application in evolutionary biology. This type of cladistic analysis ideally complements sequence-based phylogenetic reconstructions. Whereas the latter ideally allow us to make inferences about the timing of splitting events by comparing the distances between certain taxa, the former complements conventional sequence analyses from a completely different perspective.

Theoretically, a SINE/LINE analysis can be done on both the multilocus and single locus levels. The former can be done in a technically easy hybridization experiment and does not require any sequence information apart from the SINE/LINE sequence itself. As far as a SINE with a high copy number in a certain taxon to the exclusion of another is concerned, the pure presence/absence pattern can be informative about phylogenetic affiliations and unite all subtaxa belonging to the former into a monophylum. For the monolocus approach, sequence information from the region surrounding the integration is necessary to design primers that specifically amplify the orthologous regions in the taxa under scrutiny.

Molecular cladistic SINE/LINE analyses are becoming widely used since they are regarded as representing powerful noise-free Hennigian synapomorphies (Shedlock and Okada 2000).

The latter is due to the fact that SINEs/LINEs as marker systems in evolutionary biology display rates of homoplasies and reversals that can be safely assumed to be far lower than observed for sequence data and complex morphological characters. As far as the relatively unspecified targets of the LINE integration machinery (see above) and the size of a typical eutherian and primate nuclear genome are concerned, the chance of SINE/LINE integrations independently involving the same targets is negligible even over evolutionary time scales. Although, e.g., repeats may preferentially integrate into locally AT-rich regions in
chromosome R bands (Korenberg and Rykowski 1988; Matera et al. 1990) or might have a tendency to prefer integration into DNA regions that can adopt alternative structures, for instance kinks, the LINE machinery does not have a constant and unambiguous integration site that can be specified by comparing the SINE/LINE flanking direct repeats. Even if a closely spaced target site has been hit in different evolutionary lineages by an integration event, the direct repeat and the structure and characteristics of the SINE/LINE itself can give precious information on the true orthology of SINE/LINE integrations. SINE/LINE insertions, present at the orthologous positions in different genomes, can thus reliably be considered to be identical by descent.

A further huge advantage of SINE/LINE integrations compared to sequence data is the virtual absence of homoplasies resulting from character reversals. Although evidence was recently reported, after comparing complete human, chimpanzee, and rhesus monkey sequences, that in rare cases a certain SINE can be precisely reexcised, creating a situation similar to the incomplete lineage-sorting process mentioned below (van de Langenmaat et al. 2005), the rate of integration-reversals must be lower by orders of magnitude than the rate of substitutional reversals. Thus in primates, contradictory results among individual presence/absence analyses that need to be generated if reversals occur at decent frequency were simply not observed in the analysis of hundreds of different orthologous Alu-SINE integrations involving the different primate infraorders. Compared to the vast bulk of the molecular cladistic information of a certain branching event in question, a precise reversal would be either not remarked at all or clearly recognized as an autapomorphy.

Of course, major parts of chromosomes can be deleted during the course of evolution and might encompass SINE/LINE sequences. However, a molecular cladistic analysis can discriminate between the absence of an orthologous SINE/LINE and the absence of the respective stretch of DNA. In a typical experiment, PCR primers need to be placed well away from the SINE/LINE and should incorporate some single-copy flank. A shorter PCR fragment thus signals the absence of a SINE/LINE, a longer fragment its presence. A deletion of the SINE/LINE and flanking regions would not generate any PCR fragment. Thus, the presence of a SINE insertion yielding a PCR fragment that is longer than the insertion absent-fragment reflects the derived character state. An unoccupied target site or shorter PCR fragment that—depending on the age of the very integration—can be easily diagnosed by the presence of the short repeat-flanking sequence in an unduplicated version represents the ancestral character state. This in turn results in a well-defined character polarity of these molecular cladistic markers. Naturally, SINE-based molecular cladistic analyses do have some limitations that might be overcome by applying different experimental strategies.
One problem is that most SINEs are located in noncoding regions, which usually display considerable divergence when analyzing taxa that coalesce deeply. In consequence, PCR primers might not possess sufficiently conserved targets to allow efficient amplification across highly divergent taxa. Schmitz et al. (2001) developed a strategy and used intronic SINEs as markers that allow the construction of primers that target into conserved exonic regions, thus amplifying the encompassed intron together with the SINE. This strategy can even help to manage interordinal amplification problems.

Another problem, which affects every polymorphic marker system, is the uneven distribution of ancestral polymorphisms into progeny lineages. This phenomenon is termed incomplete lineage sorting and affects very recent splits as well as deeper, consecutive splitting events that took place within a short period of time. It is obvious that every derived character state is the result of a mutation (or integration) that happened in the germline of one individual of a certain taxon. This mutation will be enriched in the succeeding generations but might not reach fixation. If ancestral polymorphisms are incompletely sorted into progeny lineages, gene trees might deviate from species trees. The fixation of a neutral polymorphism (one can safely assume that many SINE/LINE integrations, e.g., in noncoding regions, do not have a functional consequence) is dependent on the effective population size. Since the effective population sizes accompanying a splitting event are mostly unknown, it is difficult to determine whether a SINE marker was fixed in the ancestral progenitor population at the base of the consecutive split. It is therefore of great importance to collect split-specific SINE information at multiple independent loci in order to test for any inconsistencies in marker interpretations that might have been caused by lineage sorting (Takahashi et al. 2001). The WGS data currently at hand, and those produced in the near future, will allow us to define such a large number of markers that it should be easy to avoid errors caused by lineage sorting.

2.2.3 Molecular evidence of primate origins

Currently, inferences based both on conventional sequence analyses and RGC support the existence of a Euarchontoglires-monophylum in that Primates and Glires are assembled together with dermopterans and scandentians. Thus, the two eutherian orders Dermoptera (flying lemurs) and Scandentia (tree shrews) are noted as potential extant sister groups to the primates, although another construction would be that both are united as sister groups, branching off before the divergence of primates. Molecular evidence—both nuclear and mitogenomic—points overwhelmingly to an exclusion of bats from the Euarchontoglires (Pumo et al. 1998;
Teeling et al. 2000). The supposed close phylogenetic relationship between Dermoptera and Chiroptera might therefore be based on convergences that emerged with the evolution of gliding or flight.

From a molecular point of view, only contradictory information is available on the position of the scandentians. This is due to the fact that hitherto only limited phylogenetic information could be gained from short alignments and somehow unbalanced data sets. Moreover, there is growing evidence that the inclusion of mtDNA sequences in the analysis of the major primate groups and primate-related eutherian orders is linked to obtaining contradictory results. Starting from sporadic observations related to the proposed close affiliation of dermopterans to anthropoid primates (Arnason et al. 2002), initially Schmitz et al. (2002a, b), Schmitz and Zischler (2003), and recently Raina et al. (2005) proposed that there are several evolutionary jumps related to the nucleotide composition of mtDNA of primates and related eutherian orders. As the cause of this, the latter proposed the asymmetric and asynchronous replication mode of mtDNA, involving two different origins for the two different DNA strands, since the replication is started at different time points. This “classical” mechanism of mtDNA-replication is probably the one that usually takes place, leaving considerable parts of the mtDNA in a single-strand configuration over longer periods during the replication. During single-strand configuration, DNA is prone to hydrolytic damage, namely deamination of cytosine and adenine, causing transitions to thymine and guanine, respectively. The resulting changes in the overall DNA composition might therefore lead to unorthodox and artifactual phylogenetic affiliations caused by similar base compositions rather than emerging from true phylogenetic information. Most unexpected to primatologists was the close affiliation of dermopterans to anthropoid primates, which was proposed from a composed mtDNA and nuclear data sets (Murphy et al. 2001a) and from complete mtDNA information alone (Arnason et al. 2002). The term “Dermosimii” was even introduced to emphasize the close phylogenetic relationship between anthropoid primates and the dermopterans. Raina et al (2005) speculate that for similar reasons the position of Tupaia in the eutherian tree cannot be correctly determined on the basis of mtDNA sequence comparisons.

Finally, the hitherto underappreciated diversity of the scandentian taxa might pose another problem for inferring the correct phylogenetic affiliations of primate-related eutherians. Olson et al. (2005) correctly state that traditional interordinal studies, both morphological and molecular, often included only a single species of tree shrew. It is conceivable that by not adequately representing scandentian diversity, phylogenetic artifacts can also be created by the molecular approaches. Molecular data now at hand favor a scandentian monophyly and a sister group relationship between Ptilocercus and the remaining scandentian
genera. The phylogenetic relation to primates remains unclarified by conventional studies of substitutional sequence evolution.

What about the information obtained from repetitive DNA? Several Euarchontoglires representatives display SINEs in their genomes that are obviously derived from 7SL RNA, an RNA polymerase III transcript that constitutes the RNA part of the signal recognition particle. These include the rodent B1, the primate Alu-SINES, and the scandentian Tu type I and II SINE families (Nishihara et al. 2002). No 7SL RNA-derived SINE family has yet been detected in rabbits or flying lemurs. The lagomorph genome harbors a SINE family designated the C repeat (Cheng et al. 1984), the latter however is presumably derived from tRNA genes (Sakamoto and Okada 1985). Schmitz and Zischler (2003) and Piskurek et al. (2003) characterized tRNA-derived elements in the flying lemur genome. These tRNA-derived elements apparently constitute the major class of SINEs in these genomes; however they cannot be traced in primate or scandentian genomes. This is corroborated by the fact that the respective paralogs display only little divergence, suggesting a coalescence to a recent ancestor. At any rate, from an alternative molecular perspective these elements separate dermopterans from the other Euarchontoglires.

As far as possible SINE information for the scandentian affiliation with other eutherians is concerned, it is interesting that both scandentian and some strepsirhine primate taxa belonging to the lorisiforms harbor a compound SINE in their genomes. The latter is composed of both a 7SL and a tRNA-derived SINE (Tu type I and II and the galago type II element, Nishihara et al. 2002). 7SL RNA is also known as the precursor of the major primate SINE the so-called Alu sequences. Primate Alu elements display a dimeric makeup that originated from a fusion of the left and right monomer (FLAM and FRAM). An even more ancient monomer unit that forms part of the Alu family was named FAM (Quentin 1992a, b, 1994). Because of the similarity of the FLAM sequence to a precursor sequence of the rodent B1 family, it can be hypothesized that the precursors of primate Alus, rodent B1, and scandentian Tu elements, the FLAMs and FAMs might have been created in an Euarchontoglires ancestor. This would imply that these monomers were transpositionally active during these times. However, these monomer sequences have not been directly observed in the genomes of tree shrews or flying lemurs, an observation which might be also explained by sequence divergence that took place in the long and independent evolutionary history of the Euarchontan orders. A substantially different initial copy number and/or activity state of these elements after speciation is another possible explanation.

Until the first WGS data on a scandentian representative are released, it needs to be stated that 7SL-derived SINEs can probably not be used for
establishing sister group relationships among the different Euarchontoglires on the monolocus level.

However, SINE-data can also be used to separate the primates from the other Euarchontoglires. This can be technically easily done on the multilocus level. Dermopterans, misplaced by mtDNA analyses, can be firmly excluded as closely related to anthropoid primates as Arnason et al. (2002) suggested. The most abundant SINE in primates, the Alu-SINE which has about a million copies in the human genome, cannot be traced in dermopterans. If, as suggested by mtDNA analysis, the Dermosimii clade were correct, a simultaneous elimination of hundreds of thousands of Alu copies from the genome of dermopterans once they had split off from the lineage leading to the ancestor of anthropoid primates would have been necessary. This scenario can be firmly excluded.

Finally, a set of three monolocus repeat markers that are shown by representatives of all primate infraorders to the exclusion of other, non-primate eutherians including the Dermoptera (Schmitz et al. 2002b, 2003) can be defined, thus corroborating primate monophyly from a retropositional site.

Figure 2.1 summarizes the molecular evidence on Euarchontoglires obtained from retroposons. It can merely be stated that scandentians, dermopterans, and primates each represent monophyletic groups. Their firm consecutive splitting needs to await more data—the anticipated Tupaia WGS sequence will be one step in the right direction.

2.2.4 Infraordinal primate phylogeny

During the last decade, the number of primate species recognized has constantly increased. Even traditional classifications (Rowe 1996) included more than 230 species in this order, but nowadays well beyond 350 different species have been proposed (Groves 2001). Intensified field work, as well as the incorporation of molecular analysis, led to this large number. We may assume that in the foreseeable future, molecular data will reveal that current species diversity estimations for the extant primates are very conservative and that morphology-based classification has underestimated the genetic diversity. Primates may be traced back to a common ancestor—several authors quote Purgatorius as a possible candidate for the primate MRCA. The time of primate origin has constantly shifted backward over the last few years. Tavaré et al. (2002) proposed a time of 85 Ma based on paleontological evidence, and molecular data also tend to shift the MRCA further back in time than formerly thought. The origin and remarkable diversification of primates, rendering it one of the most successful eutherian orders, is usually
linked to three different theories: the arboreal theory, the visual predation theory, and the angiosperm radiation theory. Several aspects of these theories are constantly disputed, among them the general acceptance of the first primate ancestors being relatively small animals (Soligo and Martin 2005), a statement for which neither paleontological evidence nor extant taxa provides evidence. In addition, molecular data on the genes involved in color vision cannot rule out that trichromatic color vision was already present in the primate stem lineage, ruling out that nocturnality in primates is an ancestral character state (Tan and Li 1999;
Jacobs and Deegan 2003). Primates evolved full stereoscopic vision, reflected in the orbital convergence of early primates. In parallel, the olfactory system was reduced, mirrored in the reduction of the snout and the brain areas involved in the processing of chemosensory information. The number of functional genes for both taste and olfactory information was reduced during primate divergence and on the lineage to the hominoids. The increase in frequency of nonfunctional pseudogenes might be due to reduced selective pressures (reviewed in Hacia 2001).

In contrast to the uncertain interordinal phylogenetic relationships of primates with other extant eutherian representatives, far less disagreement exists about the phylogenetic affiliations between and within the major extant primate groups. The molecular data, mainly SINE evidence, are discussed below.

Obviously, by just taking the pure number of possible markers into account, the most abundant primate SINES, the Alu sequences, represent an inexhaustible source of molecular cladistic markers. However, it needs to be mentioned that all other sorts of retroposons can be taken as marker systems as well. ERVs, the remnants of ancestral germline infections by retroviruses, might serve as an example (Sverdlov 2000). However, the technical ease and speed at which Alu-markers can be established from WGS data of humans, chimpanzees, and rhesus monkeys—not mentioning the taxa for which WGS data will soon be available—have brought them into center stage in the molecular cladistics of primates.

Over one million Alu sequences have been identified in the human genome, thus comprising more than 10% of its overall sequence (Li et al. 2001). Alu-SINES can be traced back to a fusion of two monomers that took place in an ancestor of the extant primates. The resulting 300 bp element colonized primate genomes exclusively, thus representing a molecular synapomorphy that separates primates from non-primate species. Reliable estimates for Alu copy numbers in other primate species need to await more WGS data, but great apes and strepsirhine representatives are estimated to harbor several hundred thousands of Alu copies in their genomes as well. Alu sequences are frequently found in noncoding regions such as introns, untranslated regions of genes, and intergenic genomic regions (Deininger and Batzer 1993; Makalowski et al. 1994).

Alu sequences harbor diagnostic mutations that permit partitioning them into different subfamilies. This is assumed to be due to the fact that only a small subset of Alu elements was retropositionally active for a certain time period during primate divergence, thus serving as an Alu-source or “master” gene (Deininger et al. 1992).

Two contrasting models of SINE evolution have been linked to the retropositional activity of offspring SINES. A few active “master” SINES that are
capable of retropositional amplification are invoked by the master gene hypothesis (Deininger et al. 1992), whereas the “multiple source gene model” predicts that the progeny SINE copies can potentially propagate to the same extent as the parent SINEs they are derived from (Schmid and Maraia 1992). Experimental and bioinformatic evidence currently favors the “sprout” or multiple source model (reviewed in Shedlock and Okada 2000; Piskurek et al. 2003).

Alu-source genes thus accumulate diagnostic new mutations that give rise to a set of Alu subfamilies retroposed in chronological order (Batzer et al. 1996). Some of the young Alu elements described for humans, chimpanzees, and baboons have been integrated at their genomic location so recently that they have not become fixed in the respective taxon (Batzer and Deininger 1991; Perna et al. 1992; Batzer et al. 1994). Interestingly, active Alus are predominantly mobilized in the germline, with a bias toward the male germline. Probably epigenetic modifications are erased for a short time period during sperm cell development, inducing an increased retropositional activity of Alu-SINEs (Hedges and Batzer 2005).

The application of a molecular cladistic SINE analysis in primate evolutionary biology was first suggested by Ryan and Dugaiczyk (1989). However, the Okada group, working on different organisms, was the progenitor of the use of SINEs in molecular phylogeny in general. A short description of the current state of SINE-related phylogenetic research within the primates is given below.

A comprehensive SINE analysis of the strepsirhines is now to hand, giving a well-supported answer to long-standing phylogenetic questions related to the deepest primate split (Roos et al. 2004). Members of the strepsirhines constitute ca. 20% of all living primate species and can be subdivided into two groups, the Malagasy Lemuriformes and the Lorids found on the two continents Africa and Asia. Distribution of extant strepsirhines and paleocontinental reconstructions suggest an African origin of strepsirhines with two subsequent migrations to Madagascar and Asia. An oceanic barrier separated Madagascar from Africa and the Indian subcontinent from at least 165 and 88 Ma, respectively; thus even accepting the oldest dates for a primate origin would require the crossing of a huge water barrier. Both the timing and the physical crossing of the Mozambique channel from Africa to Madagascar remain much disputed. A long-standing question is therefore how often Madagascar was colonized by lemuriform ancestors. Molecular cladistic data could assist in answering this question by giving an unequivocal proof of the monophyly of Malagasy lemurs, suggesting a single colonization event.

A question usually formulated as a phylogenetic problem is related to the questionable cheirogaleid-galagid affiliation, based on the intracranial blood supply shared by both the Malagasy cheirogaleids and the African galagids.
The resulting tree topology would support at least two independent colonizations of Madagascar. Another problem regarding the monophyly of Malagasy lemurs is far more complicated and concerns the position of the aye-aye (*Daubentonia madagascariensis*). MtDNA data are not able to yield conclusive evidence for the aye-aye’s position and its relation to other Malagasy lemurs. Alternative topologies have been proposed, showing the aye-aye at the base of all strepsirhines, joined with the Indriidae, or not fully resolved with respect to other strepsirhines (Arnason et al. 1998).

A comprehensive analysis of strepsirhine phylogeny applying molecular cladistic markers was presented by Roos et al. (2004; Figure 2.2). By combining both multilocus and monolocus SINE evidence, SINE integrations were positive for a monophyly of Malagasy lemurs, with *Daubentonia* being a descendant of the deepest split and ruling out a close phylogenetic affiliation of cheirogaleids and galagonids. SINE-RGCs support Lorisidae as a monophyletic sister clade to the Galagidae and corroborate a common ancestry of African and Asian lorisids. Furthermore, one SINE integration links the two diurnal or cathemeral

![Figure 2.2](image-url)

**Figure 2.2**
Phylogenetic affiliations of extant strepsirhine representatives. The tree topology was taken from Goodman et al. (2005) and the number of molecular cladistic arguments (SINE insertions as RGC) displayed above the branches they support (Roos et al. 2004). The genera *Euoticus* and *Indri* were not included in the SINE analysis (in gray). The position of *Euoticus* could not be unequivocally determined by comparative sequence analysis.
lemuriform families, the Lemuridae and Indriidae, indicating an autapomorphic reversal to nocturnality by the indriid genus *Avahi*.

On the intraordinal taxonomic level, the most striking cladistic problem concerning the phylogenetic pattern of extant primates is represented by the phylogenetic affiliations of *Tarsius* to other extant primates. Differing taxonomic categorizations—the prosimian/simian and strepsirhine/haplorhine distinctions—reflect the uncertainty concerning the tarsier’s position in the primate phylogenetic tree. Its position is not easy to determine from molecular approaches since *Tarsius* is the only surviving genus of a formerly diverse group of Eocene tarsiiforms. Thus, this one genus cannot properly reflect the complete tarsiiform diversity that once existed. Moreover, the long independent history of *Tarsius* and its high likelihood of acquiring autapomorphies might confound phylogenetic tree reconstructions.

From a molecular perspective, mtDNA data are notoriously difficult to interpret, giving rise to artifactual assemblages based on mtDNA comparisons. First, deviations from a purely neutral mode of mtDNA-evolution were invoked to explain this apparent misplacing (Andrews et al. 1998). However, the unexpected positioning of *Tarsius* in an mtDNA-based phylogenetic tree could equally be explained by a base composition plasticity of mtDNA (Schmitz et al. 2002a, b; Raina et al. 2005). In contrast to mtDNA, most nDNA data sets firmly place *Tarsius* as a sister to the anthropoid primates, thus supporting the haplorhine grouping. SINE evidence on *Tarsius* has so far been obtained by both substitutional sequence analysis of Alu-SINEs and as molecular cladistic markers. The former approach resulted in placing *Tarsius* as a sister to the anthropoid primates (Zietkiewicz et al. 1999). The latter molecular cladistic SINE approach analyzed the presence and absence pattern of 118 loci with intronic Alu sequences (Schmitz et al. 2001). Initially from these three Alu, SINEs were determined to be present at orthologous sites in *Tarsius* and all anthropoids, to the exclusion of the strepsirhines and outgroup representatives. In the latter, unoccupied integration target sites were detected. Another integration of an unfused Alu monomer was later added to the list (Schmitz and Zischler 2004). This Alu monomer might have been created by a deletion or it can be speculated that a monomer that was retropositionally active before the primate divergence regained the capacity to spread through the genome. Both the analyses of presence/absence and of substitutional evolution are not comparable, the latter representing a multilocus approach, whereas the former pinpoints one orthologous region in the primate genomes. Further retropositional evidence for a close affiliation of anthropoids and *Tarsius* was also provided by Kuryshev et al. (2001).

Taken together, all these repeat analyses result in four strong molecular cladistic arguments in favor of a sister group relationship of tarsiers and
anthropoid primates. Since in more than 100 analyzed Alu-loci no inconsistency leading to contradictory interpretations could be detected, the likelihood of confounding these results by lineage sorting or reversals is exceedingly low. Thus, monolocus evidence as well as multilocus evidence from different SINEs (Zietkiewicz et al. 1999; Kuryshev et al. 2001) corroborates the monophyly of haplorhine primates (Anthropoidea and Tarsiiformes) from an entirely new molecular perspective (Figure 2.3).

Besides a scanty fossil record, the main problem that is encountered doing phylogenetic research on New World monkeys is due to the radiation-like evolution of extant platyrrhine diversity. The huge diversification in a short time span after the New World was colonized by the proto-platyrhines often results in long branches leading to closely spaced splitting events—a problem neither morphologists nor molecular biologists can adequately deal with without a great deal of data at hand.

Figure 2.3
Haplorhine phylogenetic affiliations displaying the sister group position of Tarsius to anthropoid primates—supported by four SINE insertions—and a broad taxonomic representation of platyrrhines (topology as in Goodman et al. 2005). Molecular cladistic arguments are summarized from Schmitz et al. (2001), Schmitz and Zischler (2004), Singer et al. (2003), and Ray et al. (2005). Taxa in gray were not tested in the SINE analysis
At least 16 New World monkey genera are recognized. Both the genus and species numbers fluctuate. Several scenarios for platyrrhine origins and transoceanic migration events have been noted. The most basic issue is that of a common New World monkey origin and the question of how the platyrrhine ancestors colonized their present geographic range. WGS data for one platyrrhine representative are about to be generated that will fuel the development of marker systems that can be subsequently used in comparative approaches to elucidate New World monkey phylogeny. Initially SINE data from three independent loci were presented that strongly point to a monophyletic origin of extant New World monkeys (Singer et al. 2003) corroborating SINE evidence in the major histocompatibility complex (Kriener et al. 2001) and other molecular data based on sequence comparisons. Recently, Ray et al. (2005) added an impressive list of 183 molecular cladistic Alu-SINE markers that were informative for platyrrhine evolution. The number of molecular cladistic markers pointing to a platyrrhine monophyly now approaches 100. Several molecular cladistic SINE arguments support a sister group relation of the Cebidae and Atelidae with the Pitheciidae branching off before the common ancestor of Cebidae and Atelidae. Together with the SINE information provided by Singer et al. (2003), three orthologous SINEs represent shared-derived character states for the Cebid clade including Aotus, Saimiri, Callithrix, and Saguinus. The amount of arguments basically rules out a confounding of the resulting interpretations by lineage sorting. A comparison of SINE-based evidence to conventional sequence analysis and fossil evidence is given in more detail by Ray et al. (2005).

Catarrhini, the cercopithecoid monkeys and hominoids, represent the youngest group of primates. Extant catarrhines can be partitioned in the Cercopithecoida (Cercopithecinae and Colobinae) and the hominoids, including the gibbons, orangutans, gorillas, chimpanzees, and bonobos. Salem et al. (2003) used altogether 133 Alu-SINEs to delineate the hominid phylogeny that is congruent with previous views obtained from all sorts of data. Gibbon phylogeny is difficult. A consensus is emerging that favors recognizing the four genera Nomascus, Symphalangus, Bupithecus, and Hylobates. Their phylogenetic affiliation is disputed, but the unavailability of material suitable for molecular analyses renders thorough taxonomic sampling near impossible. As to the marker site, gibbon WGS data are in the pipeline and are a likely source of numerous future marker systems.

A synthetic view of catarrhine evolution based on neontological—including molecular—and paleontological data is at hand (Stewart and Disotell 1998). Xing et al. (2005) performed a SINE analysis by cross-species checking the presence of altogether 297 Alu-SINEs (Figure 2.4). The formerly less disputed phylogenetic
relationships among the Papionini, a group of catarrhines consisting of the genera *Macaca*, *Papio*, *Theropithecus*, *Lophocebus*, and *Mandrillus*, and *Cercocebus* are well supported by several SINEs indicative for the respective branches. Moreover, the affiliations among the Cercopithecini (comprising *Cercopithecus*, *Chlorocebus*, *Erythrocebus*, *Allenopithecus*, and *Miopithecus*), formerly intensively disputed, are also now corroborated by a multitude of SINE arguments that rule out any confounding of the phylogenetic interpretations caused by lineage sorting. Finally the question of monophyly of Asian and African colobines has also been approached by SINE analyses. As in the case of the gibbons, however, the Asian and African colobines can be firmly assembled but the taxonomic representation is not complete, leaving out such taxa as *Piliocolobus* and *Simias*. Thus colobines are somehow representative for the whole primate divergence. Though markers are at hand and are likely to increase in the near future, mainly due to the WGS and ENCODE sequencing initiatives, there are still quite a few gaps remaining before we will get a complete view of the phylogenetic affiliations of all extant primate genera.
2.3 Future directions

The most striking phylogenetic problem in primate evolution is still that of the enigmatic primate origin and affiliation to other Euarchontoglires. Considerable improvement of this situation will be obtained by obtaining *Tupaia* WGS data. These will help to establish meaningful comparisons of rodents with *Tupaia* and primate genomes. Markers will be defined that need to be tested in dermopterans and that cover the complete scandentian divergence. It can be anticipated that this will lead to a well-supported phylogenetic framework linking eutherian model systems and humans.

Looking at the extant taxa, molecular processes shaping primate genomes can thus be pinpointed on the platform of a settled phylogenetic framework. Moreover, a phylogenetic framework linking extant primates and primate-related eutherians might be helpful in formulating hypotheses in which fossils are incorporated, thus yielding information on the morphological character evolution and phylogenetic affiliations of extant and extinct primate species.

Furthermore, total evidence approaches linking molecular and, e.g., morphological information need to be put forward. Experimental data as well as bioinformatic tools need to be improved in this context. Maximum parsimony has been the method of choice for combined molecular and morphological data sets, and this is above all a problem when analyzing deep splits with molecular data. New approaches include Bayesian methods, which allow molecular and morphological data to have their own evolutionary models.

More sequence data will enable us to more reliably estimate divergence and origin times—clearly mtDNA, with its peculiar mode of evolution, needs a backup in the form of sequence data of sufficient quality and quantity.

To delineate character evolution, a better coverage of all primate taxa including strepsirrhines and tarsiers and the non-primate taxa is of the outmost importance.

Finally, with more sequence data and a phylogenetic framework at hand, the functionally important molecular changes also involving SINEs will be the future research directions in primate genetics. Interspersed repeat sequences constitute additional important forces that possibly alter gene function and regulation (Samonte and Eichler 2002). Exaptations of repetitive elements, epigenetic and posttranscriptional modifications linked to SINEs, the role of SINEs in triggering large-scale genomic changes including segmental duplications and inversions, are just a few of the effects SINEs can exert on genomes and transcriptomes, thus potentially triggering major phenotypic changes. They show that in the future retroposon analyses will be worthwhile experimental approaches for both pattern and process analyses in primate evolutionary research.
Acknowledgments

Thanks go to the members of the former Primate Genetics Group at the German Primate Center who worked on part of the topics mentioned in this review. Members of the EU-consortium INPRIMAT are gratefully acknowledged for discussions and providing comments and suggestions. Funding was received from the DFG and the EU (INPRIMAT,QLRI-CT-2002-01325).

References


Han JS, Boeke JD (2005) LINE-1 retrotransposons: modulators of quantity and quality of mammalian gene expression? Bioessays 27: 775–784


Jurka J (1997) Sequence patterns indicate an enzymatic involvement in integration of
mammalian retroposons. Proc Natl Acad Sci USA 94: 1872–1877
Quentin Y (1992b) Origin of the Alu family: a family of Alulike monomers gave birth to the
left and the right arms of the Alu elements.

Nucleic Acids Res 20: 3397–3401


Sakamoto K, Okada N (1985) Rodent type 2 Alu family, rat identifier sequence, rabbit C family, and bovine or goat 73-bp repeat may have evolved from tRNA genes. J Mol Evol 22: 134–140


by analysis of the insertion of retroposons.
3 Fossil Record of the Primates from the Paleocene to the Oligocene

D. Tab Rasmussen

Abstract

The early fossil record of primates is very rich, but it is limited to nonrandom windows in time and space based on vagueries of historical geology. Most primates known are fossil taxa and their value to studies of phylogenetics, paleoenvironments, and biogeography cannot be overstated. True primates demonstrating orbital convergence, reduced olfaction, relatively large brains, and grasping extremities are known from the earliest Eocene of North America, Europe, and Asia, suggesting a much older origin of the order. The Eocene was a primate golden age, with diverse radiations of adapoids, omomyoids, and anthropoids among others. The phylogenetic relationships of some groups remain uncertain, largely because they show mosaic patterns of features that do not conform to expectations derived from the few living primates. True strepsirhines are documented in the Eocene of Africa. Basal tarsiers are known from the Eocene of Asia. Undoubted early anthropoids appear in Africa by the Middle to Late Eocene. Proposals that Eocene Asian primates might be related to the African anthropoids remain hypothetical. By the Oligocene, anthropoids had diverged into catarrhines, platyrhines, and basal lineages now extinct. The apelike catarrhines that gave rise to monkeys, apes, and humans were present in Africa before the Oligocene–Miocene transition.

3.1 Introduction

We are lucky to have a very good fossil record of primate evolution in the Early Tertiary. Many dozens of genera are known, hundreds of species, and tens of thousands of individual specimens which in their geological contexts preserve useful information about morphology, time, and paleoenvironments (Fleagle 2000; Hartwig 2002). Of course, even with a rich fossil record, our paleontological view of primate evolution is still restricted to a relatively few number of geographic locations at any given point in time. This is because the geological processes that
generate substantial terrestrial sedimentary deposits in which primate remains may accumulate and be preserved are relatively rare, usually being associated with episodes of mountain building, aggrading coastal flood plains, deposition of large river deltas, sedimentation in rift valleys, or on a smaller scale, with local lake bed sediments, limestone cave breccias, and fissure fills. Most of the earth’s terrestrial surface is wearing away at any given time, not being buried. The primate fossil record is one of great richness in terms of overall number and quality of fossil specimens, but at the same time it is significantly limited to a relatively few, nonrandom windows in space and time.

The purpose of this chapter is to review the current fossil record of primates from the Paleocene, Eocene, and Oligocene with a special emphasis on its relationship to key phylogenetic issues in paleoanthropology. The phylogenetic value of the fossil record can hardly be overstated. Any given fossil taxon is represented by only a few anatomical systems—usually some teeth, and if lucky, entire dentitions, cranial remains, or postcranial skeletal elements—which pales in comparison to the complete anatomical, behavioral, and genetic data available on living primates. Conversely, the living primate record is limited to a relatively few taxa, a very small sample of all primates that have ever existed. Entire radiations of phylogenetically important groups are missing from the extant record (like three-premolar Old World anthropoids). And worst of all, the living primate record is restricted to an incredibly narrow slice of time—the present—which is no more important in primate evolutionary history than is any other equally thin slice of time from the past. Far more fossil primates are now known than living species, and the number of fossil taxa will continue to increase. Fossil primates show unique character combinations that would never have been inferred from living species, they provide information on evolutionary sequences, and they help tease apart problems of evolutionary convergence and parallelism. In this age in which DNA analyses are justifiably emphasized as a pivotal research tool, it is important to note that only a small minority of primate lineages will ever be known by their DNA. To incorporate genetic analyses into a broader, cohesive view of primate evolution, we will always be dependent on morphological analyses of fossils that allow us to extrapolate the genetic interpretations into the rich temporal and ecological framework of the fossil record.

The taxonomic scope of this chapter will be restricted to euprimates, those taxa that belong in a clade defined by grasping hands and feet, orbital convergence and a postorbital bar, and a petrosal auditory bulla. As Chapter 1 of this volume explored, there existed complex radiations of Paleocene and Early Eocene plesiadapiform mammals that may or may not be allied with euprimates. The recent discovery of primate grasping in the non-euprimate *Carpolesus*, which lacks orbital convergence, highlights how important the fossil record of plesiadapiforms and their allies will be untangling phylogenetic and adaptive questions about the
origins of primates (Bloch and Boyer 2002). In this chapter, I will devote most of my attention to undoubted euprimates that illuminate the basal divergence of the three major living clades—Strepsirhini, Tarsiiformes, and Anthropoidea.

### 3.2 The earliest euprimates

Mammalian taxa that are routinely considered in discussions of the geologically earliest euprimate include the following taxa: *Altiatlasius, Altanius, Donrusselia, Teilhardina*, and *Cantius*. *Altiatlasius* appears in the Paleocene of Africa, the other four appear nearly simultaneously at the base of the Eocene in the Northern Hemisphere, with representatives of one or more of these taxa present in Asia, Europe, and North America (Rose 1995; Covert 2002).

#### 3.2.1 *Altiatlasius*

*Altiatlasius* is known from a handful of isolated upper and lower teeth recovered from the Paleocene of Morocco (Sigé et al. 1990). The teeth were recovered from a small deposit consisting of clasts and mammal teeth of apparent terrestrial origin transported into estuarine sediment where they came to rest. This tiny glimpse of terrestrial micromammal fauna is our only significant view into the Paleocene of Africa. The lower molars have high, complete trigonids, and low talonids which are slightly broader than those found in many insectivores. Tooth position of the isolated specimens was interpreted in a reasonable way and yields a very primate-like dental pattern. If the interpretation of these remains is correct, then *Altiatlasius* would be the earliest euprimate known and would have several important implications to primate evolution. It would suggest that euprimates originated in Africa and they spread northwards into Asia, Europe, and Africa as the global-warming event of the Paleocene–Eocene boundary sent tropical rainforests expanding into the higher latitudes.

The problem with *Altiatlasius* is that its primate-like teeth mirror, in a general sense, the primate-like teeth long known from the Paleocene of the northern hemisphere, which have caused so much phylogenetic uncertainty, such as those of the plesiadapiforms and other dentally primate-like groups (such as apatemyids). If the Paleocene of the northern continents was full of groups that could be characterized as “dentally primate,” then we must consider the possibility that such groups occurred in Africa, as well. Unfortunately, it seems unlikely that substantial deposits of Paleocene age will be forthcoming from other parts of Africa any time soon. The discovery of any such geological contexts will be of great value and should be considered a paleoprimatological priority.
3.2.2 Altanius

This interesting mammal is from the Early Eocene of Mongolia (Rose and Krause 1984; Gingerich et al. 1991). It is known from several jaws preserving teeth and shows a generally primate-like dentition. The interesting debate that emerged soon after the description of Altanius was whether this genus was a euprimate or a carpolestid (a question that preceded the new discoveries about the limbs and cranium of Carpolestes; Chapter 1). The most distinctive feature of the lower dentition of Altanius was that it had a distinctly high, buccolingually compressed trigonid on the first molar and a high, relatively complex fourth premolar, which seemed to hint at the unique dental specializations of carpolestids. While new finds of better material, including upper jaws, seemed to resolve the debate in favor of euprimate affinities, the new anatomical knowledge linking North American Carpolestes to euprimates and the discovery of structurally primitive but geologically late carpolestids in eastern Asia makes Altanius of some interest in terms of testing ideas about a basal carpolestid-euprimate radiation.

3.2.3 Cantius

This is the earliest euprimate genus known from the triad of grasping extremities, orbital convergence with a postorbital bar, and a petrosal bulla. Thousands of fossils of this genus have been recovered that represent several species from the earliest Eocene of North America and Europe. While the geologically oldest species, C. torresi of Wyoming, is known only from dental remains (Gingerich 1986), somewhat later (but still Early Eocene) species of the genus include partial crania and limb skeletons that are indisputably euprimate (Dagosto 1993). The very dense, continuous stratigraphic sequences preserved in the Bighorn Basin, Wyoming, allow for confident extrapolation of evolving lineages of Cantius, illuminating an evolutionary pattern that involved the acquisition of larger body size, upper molar hypocones, and mesostyles. The earliest associated skeletal remains of Cantius suggest a Lepilemur-like vertical clinger and leaper, but one that was probably diurnal in activity pattern, with teeth and body size suggesting frugivory-insectivory (Rose and Walker 1985).

3.2.4 Teilhardina

This genus is found alongside Cantius in the Bighorn Basin of North America, as well as in the earliest Eocene of Europe and Asia. Recently, a cranium of the genus has been described from Asia; this represents the oldest euprimate cranium in the
fossil record, complete with postorbital bar, a moderate level of orbital convergence, a small rostrum by Eocene standards, and a relatively large braincase (Ni et al. 2004). *Teilhardina* is a small primate with sharp, perforating teeth that indicate insectivory, while the molars also have the characteristic broad talonid of euprimates (distinguishing them from insectivores) that probably reflects a significant reliance on fruit or other grindable plant foods.

The Asian species of the genus is *T. asiatica*, the European species is *T. belgica*, and North America’s earliest species is *T. brandti*, followed later in time by the better known *T. americana*. The North American species are linked very reliably, via a dense stratigraphic record, to a diverse radiation of other small-bodied North American euprimates known from crania and postcrania. Those will be discussed later.

### 3.2.5 Other primitive early euprimates

Another primate genus that occurs in the earliest Eocene of Europe is *Donrussellia*, a tiny adapoid from Belgium. Its phylogenetic relationships to other Eocene primates are not as well worked out as those of *Cantius* and *Teilhardina*, but it does seem to be related to an important radiation of European adapoids.

“Geologically oldest” does not mean “most primitive,” of course, and comparative anatomical studies have suggested that a North American genus, *Steinius*, which is of Early Eocene age, may be more primitive structurally than its earlier relatives *Teilhardina* and *Cantius* (Rose 1995). It is important to note that the dentitions of all three of these North American genera are very similar to each other, more so than any of them is to modern tarsiers, strepsirhines, or most other later groups of primates. This suggests that no matter which later primate groups we may tie these Early Eocene forms to phylogenetically, we are probably sampling near the base of a radiation of primates that moved north with the wave of advancing tropical conditions that mark the Paleocene–Eocene transition (Clyde and Gingerich 1998).

### 3.3 The basal divergence of extant primate clades

An important phylogenetic question is how these few known euprimates of the earliest Eocene and their close kin in younger Eocene deposits are related to the three major extant primate clades. The evergrowing body of genetic data seems to be converging toward a consensus that the split between today’s three major euprimate groups—Strepsirhini, Tarsiiformes, and Anthropoidea—occurred
within a very short temporal interval before a euprimate fossil record is available (Yoder et al. 1996; Yoder 2003). The term “strepsirhine” is used here in the strict sense of the clade containing today’s tooth-combed prosimians, not in the paraphyletic sense that may converge on the term “prosimian” (Rasmussen and Nekaris 1998). This basal three-way divergence may prove to be an unresolvable or difficult-to-resolve trichotomy analogous in ways to the debates several years ago about the human-chimp-gorilla clade. Most analyses involving time estimates—whether by molecular clock or other means—now suggest that the deep three-way split occurred before the beginning of the Eocene. Age estimates range from as young as the Paleocene to quite a bit older (Tavare et al. 2002).

If a pre-Eocene date is correct, then representatives of a strepsirhine lineage, a tarsier lineage, and an anthropoid lineage would be expected in the Early Eocene, were we lucky enough to sample the right geographic regions. It is also important to note that there may be other basal primate lineages, outgroups to the three extant clades or embedded within the trichotomous branching. The three extant clades are important to us because they survived to the present, not necessarily because they dominated evolutionary or ecological dynamics in the Early Tertiary.

Most work on primate phylogeny starts with a basic euprimate dichotomy into haplorhines and strepsirhines. The dichotomy may be correct or not, given the problems of interpreting the three-way split (Yoder 2003). But the fact that the correct answer about the trichotomy is currently unknown and that we have not seen toothcombs and cannot see placentas in the earliest euprimate fossil record suggests that this dichotomy should not serve as an empirical starting point. Fossil evidence should help tell us whether these taxa are useful; theoretical perceptions about suborders should not dictate our interpretation of fossils. There are so many problems with the application of “Haplorhini” to the fossil record (Rasmussen 1994) that it would be preferable (but admittedly still problematic) to start looking at local primate radiations—well defined morphologically, temporally, and geographically—and work our way up from there to major subordinal divisions. Some early euprimates may resist being affined confidently with strepsirhines, tarsiiformes, or anthropoids, due to a lack of sound, multi-character, and stratigraphically buttressed fossil evidence.

To clarify the nomenclature used here: none of the three main euprimate divisions—Tarsiiformes, Strepsirhini, or Anthropoidea—is used in a paraphyletic sense. A primate that cannot be attributed confidently to “Haplorhini” does not automatically become a strepsirhine. In this chapter, primates can hold “prosimian” status, meaning we do not know to which of the three extant branches it is closest.
3.4 The Adapidea

Phylogenetic analyses of Early Eocene primates have traditionally tied *Cantius* and *Donrussellia* into a series of very important and geographically distinct radiations of a single family Adapidae, or as more commonly construed in recent years, into a cluster of families classified together as adapoids or adapiforms (Franzen 1987). These adapoids are generally characterized as arboreal quadrupeds and leapers, with diverse diets including species specialized toward frugivory and folivory, and body sizes ranging from those of mouse lemurs to howler monkeys (Gebo 2002). They have vertical, spatulate incisors, often with fused mandibular symphyses, and in some cases, at least, sexual dimorphism in canine and body size. The cranium is characterized by relatively heavy faces with small cranial capacity, simple postorbital bars, and petrosal bullae with free ectotympanic rings like those of extant lemurs. The carotid circulation is variable, most notably with some species having large promontory canals and others lacking these.

The adapoids have often been characterized as “lemur-like,” but the extent to which this reflects primitive primate similarities or actual phylogenetic ties remains unresolved. Authors of the last few decades have not been able to make a sound case for an actual adapoid-lemuroid clade, although adapids are often classified as strepsirhines or prosimians in a paraphyletic sense. In sharp contrast, a number of researchers over the years have concluded that a small handful of derived specializations shared by adapoids and anthropoids suggest that adapoids are the prosimian-grade ancestral stock of anthropoids (see review by Miller et al. 2005). This phylogenetic notion was developed in a time before the discovery of very primitive, Eocene anthropoids in North Africa. The possible synapomorphies presented in support of this hypothesis include spatulate incisors, lacrimal foramen positioned inside the orbit, presence of a large zygomaticofacial foramen, mesial canine groove, sexual dimorphism in canine and body size, and fused mandibular symphyses.

3.4.1 Notharctines

In North America, *Cantius*, or a form similar to it, gave rise to a major radiation of diurnal, herbivorous, arboreal primates that includes some of the best-known fossil primates. This group can be classified as the subfamily Notharctinae, now usually placed within a family Notharctidae (which in most taxonomic schemes also includes Cercamoniinae). Notharctines are common components of the tropical faunas found in the Early Eocene of the Rocky Mountains.
(Wasatchian and Bridgerian Land Mammal Ages), but they are extinct in that region by the Middle Eocene (Uintan). The Rocky Mountain taxa include *Cantius, Pelycodus, Copelemur, Smilodectes*, and *Notharctus* (Figure 3.1). In the lower latitudes of southern California (which lay even more southerly during the Eocene), notharctines persisted into the Uintan (represented by one genus, *Hesperolemur*) (Gunnell 1995). *Smilodectes* and *Notharctus*, which are known from crania and nearly complete limb skeletons, were large species (lemur or monkey-sized) characterized by folivory, leaping adaptations, and sexual dimorphism. They have been reconstructed as arboreal, gregarious primates similar adaptively to certain of the larger-bodied extant strepsirhines such as lemurids and indrids.

The notharctines became extinct at the end of the Bridgerian Land Mammal Age in the Rocky Mountain region, probably in association with a gradual loss of rainforest and closed woodland habitats (Townsend 2004) and, by inference, a slide toward cooler, drier climatic conditions. They left no descendants in North America but have been linked with the Late Eocene Asian radiation of amphipithecines (Ciochon et al. 2001).

### 3.4.2 Cercamoniines

In Europe, early *Cantius* or a form like it gave rise to a tremendous Eocene radiation of primates often classified together in the subfamily Cercamoniinae.
(Gingerich 1977a; Franzen 1987, 1994; Godinot 1998). Interestingly, the cercamonine derivates of primitive *Cantius* developed a hypocone from the cingulum (as do most primates, including anthropoids) but unlike the evolutionary derivation seen in notharctines in which the hypocone gradually splits away from the protocone. The cercamoniines represent one of the most diverse primate radiations documented in the fossil record, but most taxa are individually poorly known and understudied. Body sizes ranged from taxa smaller than a mouse lemur (*Donrussellia* and *Anchomomys*) to monkey-sized (4000 g) heavy-duty frugivores with deep jaws and flat, grinding molars (*Cercamonius*). They include primates interpreted as insectivores, frugivores, and folivores. Postcranial diversity is poorly documented, but they include at least a variety of arboreal quadrupedal and leaping forms (Thalmann et al. 1989; Franzen 1994).

Geographically cercamoniines are most diverse in Europe, but remarkably, and in contrast to most other Eocene radiations, they are also known from North America, Africa, and Asia. The lone North American genus is *Mahgarita*, known from the Late Eocene of Texas, which closely resembles the European genus *Europolemur* (Figure 3.2). The cranial remains of *Mahgarita* have been the subject of analyses regarding the possible phylogenetic relationships of adapoids to either strepsirhines or anthropoids (Rasmussen 1990). The African cercamoniine taxa include *Aframomius*, surprisingly similar in dental details to both *Europolemur* and *Mahgarita* (Simons et al. 1995; Figure 3.3) along with a variety of smaller, *Anchomomys*-like forms, discussed below. The Asian cercamoniine is known from a few dental remains of Middle Eocene age (Beard et al. 1994). Cercamoniines range in age from Early Eocene to Early Oligocene, a record of *Anchomomys* being the sole primate survivor of the Grand Coupure in Europe. Cercamoniinae is a diverse, widespread Eocene clade which demands further study. Extrapolations about cercamoniine paleobiology and phylogeny derived from the better-known notharctines and adapines may be off base. Phylogenetic discussions about an adapoid derivation of Anthropoidea have more often centered around cercamoniines than any other group.

Vertebrate fossil remains from the Middle Eocene site of Messel, Germany, are among the most spectacular ever found from the Early Tertiary (Franzen 1994). Whole or partial skeletons were washed into a toxic lake that prohibited bacterial decomposition or scavenging by lake-living vertebrates. The result is beautiful, pressed remains of vertebrates, including a few of primates. One of these preserves the stomach contents of a cercamoniine, the first ever reported for an Early Tertiary primate (Franzen and Wilde 2003). The primate had been eating seeds and fruit.
3.4.3 Adapines

This group of European adapoids includes *Adapis* and *Leptadapis* which, along with the notharctines, are among the best-known adapoids. They appear in the Late Eocene Paris Basin of Europe, after most cercamoniines lineages have disappeared. The earliest record of an adapine is from the Middle Eocene of Asia (Beard et al. 1994), and it now seems likely that they are a group that spread into Europe late, after an origin in the Far East. *Adapis parisiensis*, an animal about the size of a slow loris (*Nycticebus*), shows skeletal features indicating a
Figure 3.3
Mandibles of *Aframonius diedes*, a cercamoniine adapoid from the Late Eocene of North Africa. This is the largest primate known from the diverse community of quarry L-41, Fayum, Egypt; other primate taxa preserved at that quarry include mouse-lemur-sized prosimians (*Anchomomys, Wadilemur*), a possible aberrant strepsirhine (loris-sized *Plesio Pithecus*), several callitrichine-sized three-molared anthropoids (*Proteopithecus, Arsinoea, Serapia, Qatrania*), and a two-premolared catarrhine (*Catopithecus*). This higher-level taxonomic diversity hints at Africa’s important role in early primate diversification, although a shortage of productive Eocene localities on that continent stymies efforts to understand the primate radiation there. (From Simons et al. 1994; illustration by B. Miljour)
quadrupedal and slow-climbing mode of locomotion, and its teeth suggest specialized folivory. *Leptadapis* was a larger animal (the size of a howler monkey, *Alouatta*), also folivorous, with a more generalized, quadrupedal limb skeleton (Dagosto 1983). Both taxa show sexual dimorphism (Gingerich 1981). Phylogenetically, adapines seem not to have given rise to any later groups, although a specific tie to strepsirhines has been considered at times (Gingerich 1975).

### 3.4.4 Amphipithecines

For many years, a few poorly preserved teeth and jaws of enigmatic primates known as *Amphipithecus* and *Pondaungia* were known from the Middle to Late Eocene of southern Asia, with uncertainty being expressed over whether these were adapoids or anthropoids (Ciochon and Gunnell 2002). In recent years, intense research interest in a possible Asian origin of anthropoids has spurred significant new collecting efforts that have yielded an important new sample of material belonging to these two genera and newly described relatives (*Siamopithecus*, *Myanmarpithecus*) (Ducrocq 1998; Takai et al. 2001). The new remains of cranial fragments and postcranial limb elements seem to establish solidly that the amphipithecines are very adapoid-like (Ciochon et al. 2001; Gunnell et al. 2002). Certain postcranial and dental features, including the presence of a pseudohypocone and mesiodistal compression of the paraconid and metaconid, are distinct similarities to notharctines, although the two clades show plenty of differences as well. The paleobiology of the amphipithecines, putting aside controversial phylogenetic issues, has been expertly evaluated (Kay et al. 2004).

Ironically, the anatomical features that initially generated interest in amphipithecines as possibly being anthropoids—heavy, deep jaws with broad, low-crowned, thick-enamed molars—have in the last 20 years been disproven as characters of basal anthropoids. Amphipithecines have persisted as candidates for anthropoid ancestry because they do resemble later Oligocene anthropoids from Africa, and also because of the attention directed to a very different group of small-bodied, insectivorous primates that co-occur with amphipithecines in Asia, the eosimiids (see below). At this point, a specific cladistic tie between amphipithecines and anthropoids seems remote unless they are both part of a broader adapoid–anthropoid radiation (which is a distinctly minority view). Apart from their geographic address, there are essentially no characters that meaningfully link amphipithecines and eosimiids to each other, which would seem to be the critical phylogenetic question at hand in assessing the only two known Asian groups of a hypothetical basal radiation of cladistic anthropoids.
3.4.5 Sivaladapines

In southern Asia, adapoids survived the global-cooling event marking the Eocene–Oligocene transition, which spelled extinction for most northern continent primates. Adapoids crop up in outcrops of Late Miocene age in India and Pakistan, and it is a shame that we only missed them by a few million years because their DNA and soft tissue structures might well have earned them status as phylogenetic enigmas equal to tarsiers. As it is, we have jaws and teeth of large-bodied (up to 4 kg in mass), highly folivorous adapoids that moved through tropical rainforests and woodland alongside apes, monkeys, and lorisines (Gingerich and Sahni 1979, 1984) (Figure 3.4). The earliest representatives of this group appear in the Eocene of China, in the form of Hoanghonius and a few other taxa (Qi and Beard 1998; Tong et al. 1999). It seems that adapines, cercamoniines, and sivaladapines were present in Asia during the Early to Middle Eocene, and if amphipithecines can be added to the adapoid radiation there, then clearly Asia was the center of higher level adapoid diversity.

3.5 Omomyoidea

Unlike the phylogenetic ambiguity surrounding adapoids, the phylogenetic position of omomyoids seems to be indisputably tarsiiform (Gunnell and Rose 2002). While not matching modern Tarsius in the extreme expression of some of its bizarre-specialized features, omomyoids do trend toward distinctly tarsiiform morphology in most cases. For example, one of the better-known taxa—Late Eocene Necrolemur of the Paris Basin—shares with tarsiers styliform incisors, guttered glenoid fossae for the temporomandibular joint, a tubular ectotympanic, the presence of a suprameatal foramen, greatly elongated navicular and calcaneus, and a fibula fused to the tibia (Rosenberger 1985). Putative synapomorphies show up in the teeth and masticatory system, the basicranium, and the locomotor skeleton, providing support for the idea that these are not isolated cases of convergence. It could be argued that the styliform anterior dentition is primitive for primates, in which case the spatulate incisors of adapoids become possible synapomorphies linking them with anthropoids (Gingerich 1977b; Rasmussen et al. 1995; Rose 1995).

One of the key corollaries of the hypothesized phylogenetic tie between Early Tertiary omomyoids and Tarsius is that the partial postorbital closure of Tarsius must have been derived independently from the postorbital plate of anthropoids because it is absent from Eocene forms. This is one of two primary lines of evidence to suggest that the proposed synapomorphy of Tarsius and
anthropoid orbits is incorrect (the other being that the *Tarsius* plates are formed of expansions of the maxilla and sphenoid, while the anthropoid plate is primarily zygomatic; Simons and Rasmussen 1989). This case is a good example of how morphological evidence preserved among the much more diverse primates of the past, which show novel character combinations, may have a critical bearing, perhaps a definitive one, on resolving controversial phylogenetic problems.

Figure 3.4
Maxilla with P3-M3 and mandible with an almost complete lower dentition of *Sivaladapis nagrii*, a monkey-sized folivorous adapoid from the Late Miocene of Asia. By this time, most Eocene adapoids and omomyoids had long been extinct, but the sivaladapines continued to thrive as part of a primate fauna that also included hominoids, cercopithecoids, and lorisines. (From Gingerich and Sahni 1984; illustration by B. Miljour)
3.5.1 Anaptomorphines

In North America, primates similar to Early Eocene *Teilhardina* gave rise to a diverse radiation of small-bodied, leaping, primarily insectivorous primates. Studies of these fossil anaptomorphines in a stratigraphic context have been very important in understanding tempo and mode in evolution and made important contributions to the punctuated equilibria debates of a quarter century ago (Rose and Bown 1984; Bown and Rose 1987). These are the kinds of phylogenetic studies that actually relate to important evolutionary issues rather than to nomenclatural tiffs. Anaptomorphine crania are known for *Teilhardina*, *Tetonius*, and *Shoshonius*. The cranium of *Tetonius* was found in the nineteenth century and has been important in shaping historical views of primate phylogeny (Figure 3.5). *Tetonius* and *Shoshonius* have each been interpreted as being

![Figure 3.5](image)

Figure 3.5
Cranium of the Early Eocene North America anaptomorphine *Tetonius* in side view (left) and the Late Eocene European microchoerine *Necrolemur* in front view (right). These two taxa were historically important in demonstrating that uniquely tarsiiform features were present in Eocene omomyid primates. The presence of tarsiiform omomyids in the earliest Eocene is consistent with molecular data suggesting the divergence of the three major extant lineages—tarsiiforms, strepsirhines, and anthropoids—occurred more than 55 Ma

within the tarsiiform clade and, therefore, outside any potential ancestry of Anthropoidea (Beard et al. 1991). Anaptomorphines were a common and diverse component of Early Tertiary communities in North America. The last surviving member of the group was *Trogolemur*, which appears as late as the Middle Eocene, after notharctines have disappeared from the Rocky Mountain region.
3.5.2 Omomyines

Ecologically, the omomyines were to North American tropical communities what the cercamoniine adapoids were in Europe—an ecomorphologically diverse radiation of primates that spanned the dietary spectrum from insectivores to frugivores and folivores, and the body mass spectrum from animals less than 200 g to some the size of lemurs (>2500 g). Also like the cercamoniines, omomyines have a broad geographic distribution; *Asiomomys* and *Macrotarsius* are found in the Middle Eocene of China (Beard et al. 1994). Omomyines probably originated from an anaptomorphine-like primate, perhaps something like *Steinius*. The radiation peaked during Bridger times (Gunnell and Rose 2002; Muldoon and Gunnell 2002). Among the best-known species is *Omomys carteri*, which is represented by hundreds of dental specimens, and an impressive collection of postcranial remains recovered from a single paleosol in the Bridger Basin (Anemone and Covert 2000). *Omomys* was a small, insectivore–frugivore with an arboreal quadrupedal locomotor pattern capable of leaping; it is interpreted as being more of a generalized, cheirogaleid-like primate than extremely tarsier-like. Another omomyine, *Hemiacodon*, also lacks some tarsier specializations such as fused tibia–fibula, although the navicular is elongated compared to those of anthropoids and strepsirhines (Dagosto 1993).

Upon the extinction of notharctines in the North American interior, several omomyines evolved large size (comparable to modern lemurs) and teeth that indicate frugivory and folivory (*Ourayia, Rooneyia, Macrotarsius*). The hindlimb skeleton of *Ourayia* indicates it retained from its omomyine ancestors certain leaping adaptations that seem to be more specialized than what is seen in cheirogaleids (Dunn et al. in press). The cranium of a Late Eocene form, *Rooneyia*, has enigmatically defied unambiguous placement in known clades. Unlike the cranium of *Necrolemur*, it is neither particularly tarsier-like nor does it resemble in any significant detail-specific features of strepsirhines or anthropoids. *Rooneyia* is probably the most important taxon to challenge the idea that all omomyoids are components of a monophyletic Tarsiiformes.

3.5.3 Microchoerines

The Late Eocene of Europe is characterized by three microchoerine genera: the small *Pseudoloris*, which resembles tarsiers in dental structure; the large *Microchoerus*, which has more cuspate teeth suggesting frugivory; and the medium-sized *Necrolemur*, the one taxon known from crania and some limb elements, which looks convincingly tarsier-like in several anatomical systems (Section 3.5;
A geologically older omomyoid named *Nannopithecus* from the Middle Eocene of Europe may be a connecting phylogenetic link between Early Eocene *Teilhardina* and the Late Eocene microchoerines; dentally it resembles anaptomorphines in many primitive respects.

### 3.6 Eosimiidae and Tarsiidae

Several jaws, teeth, and possibly referrable tarsal bones from the Middle Eocene of China have been argued to be basal anthropoids. The taxa include *Eosimias* and *Phenacopithecus* (Beard and Wang 2004). From the Late Eocene of Myanmar (alongside those enigmatic amphipithecines) comes a larger form called *Bahinia* (Gebo et al. 2002). All three taxa are placed in the family Eosimiidae, which is proposed to be the sister taxon of Anthropoidea (Beard et al. 1994, 1996; Beard 2002).

#### 3.6.1 Dental data and tarsier theory

The original dentally based hypothesis that *Eosimias* is related to anthropoid ancestry relied on proposed synapomorphies that were not particularly compelling, traits like premolar exaenodonty (i.e., buccally bulging crowns) and a 2-1-3-3 lower dental formula, which can be found in other primate groups (Beard et al. 1994; Gunnell and Miller 2001). Given the historical record showing that bold phylogenetic hypotheses born of incomplete dental evidence often turn bad (*Ramapithecus* springs to mind), why has the eosimiid-as-anthropoid hypothesis garnered so much attention? Because it is generated from the very compelling theory that anthropoids must have been derived from a tarsier-like ancestor (Kay et al. 1997; Ross et al. 1998).

The history of this idea is central to the last century of paleoprimatology (Rasmussen 1994). In the 1890s, a Dutch anatomist named Hubrecht published work suggesting that tarsiers were more closely related to anthropoids than either group was to strepsirhines based on placental structure. Later, Pocock gave the phylogenetic hypothesis a name in the Linnean hierarchy: Haplorhini. Pocock’s work was based on the comparative anatomical study of soft tissues of the nose and the retina. British anthropologists were so taken by tarsiers that the prominent scholar Frederich Wood Jones went so far as to propose that *Homo* evolved from tarsiers, not from apes. In the mid-century, paleontologists were comfortable recognizing a clade formed of tarsiers and anthropoids among living taxa, but they also were confident about an omomyid-tarsier clade nested within
that one. That idea was challenged in the 1970s by the proposal that tarsiers and anthropoids shared a common ancestor not just among the living taxa but also relative to all known fossil taxa at the time. The hypothesis has been called the “strict tarsier-anthropoid clade.” This cleaved tarsiers off Necrolemur and other Eocene tarsiiforms and attached them tightly to Anthropoidea, based on postorbital closure and middle ear configurations of carotid circulation and septum development that were not present in those otherwise seemingly tarsiiform Eocene primates. The apparent soft tissue and genetic similarities between tarsiers and anthropoids were entirely irrelevant to testing the strict tarsier-anthropoid clade because fossil omomyoids (or adapoids or any other group) could not be tested for those traits. The strict hypothesis was a uniquely skeletal hypothesis that was clearly intended to separate anthropoids and tarsiers from the Eocene radiations of omomyoids, which for all practical purposes have no noses, placentas, or DNA.

This century-long tradition projects an unambiguous phylogenetic prediction: the earliest anthropoids evolved from a tarsier-like primate that was not too tarsier-like (too specialized) to give rise to anthropoids (i.e., no long foot bones, no guttered temporomandibular joint, no styliform incisors, no fused tibiofibula, etc.). Eosimias is compelling because it is a very tarsier-like primate that is not known to have anatomical specializations excluding it from anthropoid ancestry. This is not to say that Eosimias is generalized enough to give rise to anthropoids; at this point, it merely means that we do not know whether that is the case or not, based on lack of the most salient cranial and postcranial traits. The eosimiid hypothesis of anthropoid origins will succeed or not based on future fossil finds of more complete skeletal remains.

On the basis of the available dental material of eosimiids, which includes complete lower dentitions, and a fair number of upper teeth, it appears that they are very tarsier-like animals. Similarities in the lower dentitions of Eosimias, Tarsius, and an African Oligocene primate named Afrotarsius are pervasive (Simons 2003). In the molars, these genera differ from each other mainly in the angle of the paracristid. Eosimias differs from Tarsius in that the former retains the primitive two pairs of primate incisors, while Tarsius has only one pair. In both cases, the incisors are small, fairly vertical in placement, and pointed at the apex. On this evidence, Eosimias would seem to be a good candidate as two-incisored tarsiid.

Other teeth from China have been described as members of the modern family Tarsiidae (Tarsius eocaenus, Xanthorhysis tabrumi). Paralleling these finds, a primate lower jaw and teeth that closely resembled living Tarsius was recovered from Early Oligocene sediments in the Fayum, Egypt; this was identified as a tarsiid and named Afrotarsius (Simons and Bown 1985). Later, when eosimiids
were discovered, *Afrotarsius* and *Eosimias* were quickly reclassified into a single family, which supposedly consisted of basal anthropoids (Kay et al. 1997). When a fused primate tibia and fibula consistent in size with what would be expected of *Afrotarsius* was found in the same Fayum quarry that yielded *Afrotarsius*, a quarry with no other known mammal that could accommodate such a bone, it was suggested that therefore eosimiids were tarsiid after all (Rasmussen et al. 1998). Proponents of the eosimiids-as-anthropoids idea merely dropped *Afrotarsius* from their clade, now ignoring the evidence they had used to affix it there in the first place, which is exactly the same kind of evidence that supported the *Eosimias*-anthropoid clade to begin with.

### 3.6.2 The tiny foot bones of China

From the same fissure-fill deposits that have yielded dental remains of *Eosimias*, tarsiids, cercamoniines, adapines, omomyines, and a range of non-primate mammals (Beard et al. 1994) come small isolated tarsal bones that have been interpreted as those of haplorhine primates (Gebo et al. 2000a, b). Some of these bones would represent the smallest primates ever known. Quite a few tarsals and teeth are known, and deciding on associations seems to be difficult, given the diversity of dental taxa. These bones clearly have great promise in understanding primate phylogeny and paleobiology, but at present, drawing confident inferences from them is as problematic as drawing inferences from isolated teeth. The supposed haplorhine features require confirmation by quantitative comparisons among Eocene primates, and the dental associations must be established. While these primates offer great promise, it is fair to say that there is still no confidently reconstructed taxon of Chinese fissure-fill primates.

### 3.7 True anthropoids

One of the most exciting developments in paleoanthropology in the last two decades has been the total transformation of our views regarding early anthropoids (Dagosto 2002). The traditional and widely accepted theoretical views about the adaptive origin of anthropoids centered around the globally widespread cooling and drying that occurred at the Eocene–Oligocene transition (Cachel 1979). As theory had it, the change in climate required the evolution of larger body size and a reliance on tougher, more resistant food items, which in turn required heavier jaws, robust bunodont molars, and fusion of the mandible and frontal bone. Under this adaptive perspective, it did not really matter at the taxonomic level whether early anthropoids were derived from Eocene tarsiiform primates,
or from more generalized and primitive adapoids. The scenario about an ecological switch held sway, regardless of the phylogenetic origin.

With the discovery of Eocene anthropoids in Egypt, Oman, and Algeria, all components of this picture changed (Kappelman et al. 1992; Godinot 1994; Simons and Rasmussen 1995; Thomas et al. 1999). These new anthropoids have modern-type postorbital closure but no mandibular fusion (Simons 1990; Simons and Rasmussen 1996). The body masses of these early anthropoids are tiny, the size of marmosets. The teeth are of prosimian grade, with three intact trigonid cusps raised well above the talonid basin. Hypocones are often small and weak. The anthropoids occur in paleoenvironments that show no evidence of a switch to cooler, drier conditions (Rasmussen et al. 1992). Indeed, basal anthropoids lived alongside cercamoniine adapids that are closely allied with those of the European Eocene. No dramatic ecomorphological shift is evident. Eocene anthropoids were small, insect- and fruit-eating primates with quadrupedal arboreal locomotion, differing from other Eocene primates by inexplicably having complete postorbital closure (see adaptive proposals on this by Ross 1996, 2000). It now appears possible that the earliest anthropoids represent an ancient radiation of primates in Africa independent of the adapoid-omomyoid-tarsioid radiations known by fossil evidence from the northern continents (Miller et al. 2005). Most recently, dental and maxillary remains of putative anthropoids as old as the Middle Eocene have been found in the lowest strata of the Fayum, Egypt (Seiffert et al. 2005a). One of these new primates shows evidence of greatly enlarged, nocturnal orbits in combination with teeth decidedly unlike those of tarsioids, which promises interesting new debates about anthropoid origins in the coming years.

Postorbital closure in Eocene anthropoids is, to date, documented only in two genera, *Catopithecus* and *Proteopithecus*, both from a single, important quarry in Late Eocene levels of the Fayum, Egypt. By extension, similar postorbital closure is inferred for a variety of small-bodied primates known only by teeth and jaws. I use the term “basal anthropoid” for primates with postorbital closure but of uncertain cladistic relationship to the platyrrhine–catarrhine divergence. I do not include the various, controversial, open-socketed groups that are hypothesized to be sister taxa of the closed-socket clade (cf. Beard 2002).

### 3.7.1 Basal anthropoids

*Proteopithecus sylviae* is known by a complete cranium, several jaws with teeth, and some limb bones (Miller and Simons 1997; Simons 1997; Simons and Seiffert 1999; Seiffert et al. 2000). It occurs at quarry L-41 of the Jebel Qatrani Formation.
which, along with the Messel, is one of the most remarkable vertebrate quarries of the Early Tertiary. An astonishing density of mammal, bird, reptile, and fish remains is preserved there, in a very fine-grained sediment that was deposited with minimal depositional energy (as indicated by the intact marmoset-sized crania and fragile bird bones). Unfortunately, postdepositional processes crushed the bones nearly flat in most cases, although the teeth retain their three-dimensional shape. Unlike the Messel lake beds, where articulated skeletal associations are often found, the material from quarry L-41 is disarticulated, suggesting that biotic factors, like fish feeding and bacterial decomposition, were contributing to the dispersal of elements.

*Proteopithecus* is a pivotal taxon in the interpretation of early anthropoid phylogeny. The postorbital closure of this primate is well documented (Simons 1997). Unlike all extant Old World anthropoids, it preserves a primitive dental formula of 2-1-3-3, retained today by most New World monkeys. The upper molars show a primitive trigon, relatively bunodont cusps, and a distinct hypocone rising from the distolingual cingulum. *Proteopithecus* has been interpreted as being generalized enough to have given rise to platyrrhines (Kay and Williams 1994). *Proteopithecus* will probably continue to serve as an empirical foil against competing ideas about the basal anthropoid radiations for sometime to come.

Other three-premolaried anthropoids from the Fayum’s Eocene levels include *Arsinoea*, an enigmatic taxon known only from a single jaw, and *Serapia*, an extremely bunodont species that may be related to an otherwise later radiation of primates called parapithecids (see below). An undoubted member of the parapithecid clade, *Qatrania*, is also known from quarry L-41 (Simons et al. 2001). Together, these Fayum three-premolaried primates provide evidence for a diverse radiation of basal, pre-catarrhine anthropoids in Africa. Additional finds from elsewhere in Africa may also belong to this radiation. *Algeripithecus* and *Tabelia* are tiny primates known by isolated teeth from Algeria, possibly of Middle Eocene age (Godinot 1994).

The best-known representatives of three-premolaried Old World anthropoids are Early Oligocene species of the Parapithecidae, known from the upper stratigraphic levels of the Fayum, Egypt. *Apidium phiomense* is a *Saimiri*-sized frugivore that employed arboreal leaping and quadrupedalism. It shows full postorbital closure, fusion of the mandibular symphysis, fused metopic sutures of the frontal bone, large promontory canals of the middle ear, and other distinctly anthropoid-like traits (Fleagle and Kay 1987; Simons 1995). Parapithecids differ from modern anthropoids in having very large olfactory bulbs and relatively small brains. The teeth of this group present various odd specializations. *Apidium* has a central conid on the lower molars and very low molar relief. Its close relative *Parapithecus* lacked adult lower incisors, a situation unique among
primates. The two adult canines erupted side-by-side up through the alveoli of the deciduous incisors. A more recently described cranium of Parapithecus indicates that it did retain adult upper incisors and that, in most respects, it resembles Apidium in cranial morphology (Simons 2001). Postcranially, Apidium is characterized by an extensive ligamentous joint between the tibia and fibula, which suggests a reliance on leaping, but not direct derivation from a fused tibia-fibula such as that of tarsiids. At the same time, the tarsal bones are short, like those of later anthropoids, not elongated like those of leaping omomyids, bushbabies, tarsiers.

The phylogenetic implications of the odd assemblage of character states in Apidium have been thoroughly explored (Fleagle and Kay 1987). Interpretations that preceded the discovery of the Eocene anthropoids typically placed parapithecids as the sister group of a combined catarrhine-platyrrhine clade, but the precise branching relationships among basal anthropoid lineages has become murkier now, although precise hypotheses have been put forth (Ross et al. 1998). All known anthropoid lineages are ancestrally more primitive than what would have been interpreted from parsimony analyses, indicating very high rates of parallelism among lineages, confounding use of parsimony analyses, which will take primitive members of lineages and cast them out as sister taxa.

### 3.7.2 Early catarrhines

Nowhere is the parallelism problem more evident than in the interpretation of two-premolared anthropoids from the Late Eocene. While having a catarrhine dental formula and postorbital closure, these primates are exceedingly primitive in other respects, lacking even a fused mandibular symphysis. One interpretation of this is that the two-premolared condition evolved in more than one basal anthropoid lineage (Ross et al. 1998). A potential problem with this interpretation is that there is considerable continuity up through the Fayum section that documents the acquisition of catarrhine dental and postcranial traits (Rasmussen et al. 1998; Seiffert et al. 2000; Seiffert and Simons 2001). Another possible interpretation is that the true catarrhine lineage really did originate at a point before the acquisition of mandibular fusion and other putative anthropoid traits and that parallelism between platyrrhines and catarrhines is rampant.

The putative catarrhine taxa in question, from primitive to derived, are: Catopithecus, Oligopithecus, Moeripithecus, Propliopithecus, and Aegyptopithecus. The first three are also known by unique species found in coastal deposits in Oman that lie near the Eocene–Oligocene boundary (Gheerbrant et al. 1995;
Thomas et al. 1999). The latter two, known from the Early Oligocene of Egypt, are widely accepted as basal catarrhines. *Catopithecus* is known from several crania and a few limb elements (Simons and Rasmussen 1996; Seiffert and Simons 2001) (*Figure 3.6*) and was a small, callitrichine-sized primate. The first lower molar preserves a paraconid; the upper molars have slight hypocones. The cheek teeth have a decidedly prosimian cast about them and in fact, before crania of *Catopithecus* were found, a long-standing debate concerned whether the teeth of related *Oligopithecus* were prosimian or

![Figure 3.6](image)

The Late Eocene catarrhine *Catopithecus* is represented by several crushed crania. The lower image shows one of the least crushed fragments, a lower face with all the upper teeth intact. Note the large premaxilla, the deep face, and the catarrhine dental formula. The incisors are not naturally so procumbent but have been distorted by crushing. The top image is a reconstruction featuring the postorbital region, which can be examined at various angles on several of the crushed specimens, indicating that *Catopithecus* had modern-type anthropoidean eye sockets. (From Simons and Rasmussen 1996)
anthropoid. The canines of *Catopithecus* are proportionally large in diameter and they project beyond the occlusal plane. The incisors are small and spatulate. There is sexual dimorphism in body size and canine size. The humerus primitively retains an entepicondylar foramen but displays catarrhine-like translation of the distal articular surfaces (Seiffert et al. 2000). *Catopithecus* is the best known of all Eocene anthropoids.

At the other end of the morphological and stratigraphic spectrum is *Aegyptopithecus*, a larger-bodied, heavily built frugivore with fused mandibular symphyses, loss of paraconids, large cuspathe hypocones, and bunodont, apelike molar features (Simons 1987; Rasmussen 2002). The limbs are robust and have a prosimian aspect to them in some ways (Fleagle and Simons 1982) (Figure 3.7). Unlike modern catarrhines, *Aegyptopithecus* lacks a tubular ectotympanic, retains an entepicondylar foramen, and has a very small, prosimian-sized brain. Despite these primitive features, which probably require that the derived character states evolved in parallel in platyrrhines and catarrhines, most researchers readily accept *Aegyptopithecus* as a catarrhine postdating the divergence from platyrrhines. Most
balk, however, at accepting the much higher rates of parallelism required if *Catopithecus* is a catarrhine.

A final Oligocene catarrhine is *Kamoyapithecus* from the Late Oligocene of northern Kenya (Leakey et al. 1995). The dental specimens on which this taxon is based are not well preserved and have suffered significant occlusal wear. While in many ways they resemble, in a general sense, the typical Miocene ape molar pattern, they do display an occlusal outline that is short mesiodistally and broad buccolingually, a shape also typical of the Early Oligocene *Propliopithecus* and *Aegyptopithecus*.

### 3.7.3 Early platyrrhines

The only fossil primates of Oligocene age known from South America come from the Salla beds of Bolivia. These deposits, which have yielded a small handful of primate teeth and jaws, fall chronologically in the Deseadan, or latest Oligocene, probably older than 25 Ma (MacFadden et al. 1985; Naeser et al. 1987). The morphological remains of the Late Oligocene platyrrhines can be evaluated and compared in any number of ways, but the single most important fact to come out of them is that primates were present in the isolated South American continent at that time. Their origin and mode of transport to South America remains a mystery, despite a lively literature arguing specificity of particular rafting routes on one or another ocean current, or stepping stones along one or another speculative island chain.

The morphology of the teeth of *Branisella*, the best-known Deseadan primate (and depending on taxonomic argument, perhaps the only one) remains fairly enigmatic. It shows resemblances to some modern clades of platyrrhines, but those similarities may be primitive. It also shows a remarkable resemblance in certain ways to the teeth of *Proteopithecus* of the Fayum’s Late Eocene beds (Takai et al. 2000). Whether the shared attributes provide evidence for a specialized clade containing *Branisella, Proteopithecus*, and perhaps a modern group like callitrichines, or whether the resemblances are primitive for basal, closed-septum anthropoids, is an interesting subject requiring further research. Probably the most conservative current interpretation would be that *Branisella* reflects the morphology of Eocene *Proteopithecus* because they were recently derived from a common ancestor, in which case their shared attributes might provide a fair amount of information about the basal anthropoid radiation. At the same time, they differ from each other in age by 10 Ma or so, and in geography by half a planet or so, which might generate concern that any connect-the-dots mentality is too enthusiastic.
3.8 True strepsirhines

Modern lemurs, lorises, bushbabies, sifakas, mouse lemurs, and aye-ayes are not just a bunch of leftover prosimians but are a cohesive clade that shared a common ancestor relative to all other known living primates (Yoder et al. 1996; Yoder 1997; Rasmussen and Nekaris 1998). We have more confidence in this clade than we do in Haplorhini. Finding true strepsirhines in the Early Tertiary record has proven to be difficult. Most Eocene taxa classified as strepsirhines are placed there in a paraphyletic sense.

Since the last general view of the fossil evidence pertaining to strepsirhine origins (Rasmussen and Nekaris 1998), several important finds have been made. The most important of these is the discovery of two new primate genera, Kanisia and Saharagalago, from Middle to Late Eocene beds in the Fayum, Egypt, at stratigraphic levels significantly older than quarry L-41 (Martin 2003; Seiffert et al. 2003). Both are represented by a number of isolated teeth. One tooth allocated to Kanisia is the canine of a tooth comb, complete with fine striations caused by grooming hair. This is the earliest fossil record of the tooth comb. Saharagalago is represented by cheek teeth that are remarkably similar to those of modern galagines, which has led to the suggestion that the lorisine-galagine split had already occurred by this date in the Eocene (Seiffert et al. 2003). Younger fossils from the Fayum representing the taxon Wadilemur have also been interpreted as lorisoid (Seiffert et al. 2005b). The new finds are very important for confirming that a true strepsirhine clade was present in Africa and evolving in parallel to “lemur-like” adapoids of the northern continents.

A peculiar primate from quarry L-41 has been identified as an aberrant offshoot of the strepsirhines: Plesiopithecus teras, known by jaws and a crushed cranium (Simons and Rasmussen 1994) (Figure 3.8). The primate has very loris-like lower molars and premolars, and a large, elongated, procumbent anterior tooth that may be a canine. This tooth resembles the canine of the tooth comb but is much larger, and there is no trace of the other anterior teeth (the tooth may be an incisor with the canines missing). The discovery of Kanisia and Saharagalago proves that there is great time depth in the African strepsirhine record, enough to evolutionarily generate a form like Plesiopithecus.

Other African primates that may be strepsirhines include an assemblage of small forms that are cheirogaleid like in their known cheek teeth. They also resemble small adapoids such as European Anchomomys. A detailed comparison of cheirogaleids and anchomomyins in the context of these tiny African primates is provided by Gheerbrant et al. (1993). The taxa in question include Omanodon and Shizaradon from near the Eocene–Oligocene boundary in Oman and a primate referred to Anchomomys from Late Eocene of the Fayum. Whether these prove to be more closely allied with adapoids or with true strepsirhines,
they document that at the time there did exist primates that were ecomorphologically similar to cheirogaleids in African mammal communities.

A similar, small-bodied, cheirogaleid-like primate has been found in the Eocene of Pakistan (Marivaux et al. 2001). While the researchers who described that material placed the new taxon, *Bugtilemur*, within the lemuriform clade close to cheirogaleids, it seems more likely that affinities with the contemporary tiny African primates should be considered. Our understanding of all of these vexing, little mouse-lemur-like taxa would obviously benefit from additional material. As in the case of eosimiids, only future finds can really cement the place of these taxa into a confident phylogenetic framework.

**Acknowledgments**

I thank the editors for inviting me to contribute to this volume. I also gratefully acknowledge the colleagues who have sent me casts of important fossils that I would otherwise not get the chance to examine, with special thanks to Gregg Gunnell, Chris Beard, Emmanuel Gheerbrant, and Prithijit Chatrath. I am...
indebted to Elwyn Simons and Prithijit Chatrath for always sharing with me a few of their annual new discoveries in the Fayum and for their invitations to be involved in the Fayum research. For practical help with the manuscript and the figures, I thank Hilary Brazeal.

References


Dunn R, Sybalsky JM, Conroy GC, Rasmussen DT (in press) Hindlimb adaptations in *Ourayia* and *Chipetaia*, relatively large-bodied omomyine primates from the middle Eocene of Utah. Am J Phys Anthropol


primate fossil record. Cambridge University, Cambridge UK, pp 1–4
Rosenberger AL (1985) In favor of the necrolemur-tarsier hypothesis. Folia Primatol 45: 179–194
Seiffert ER, Simons EL (2001) Astragal morphology of late Eocene anthropoids from the Fayum Depression (Egypt) and the origin of catarrhine primates. J Hum Evol 41: 577–605


4 Fossil Record of Miocene Hominoids

David R. Begun

Abstract

Hominoids, or taxa identified as hominoids, are known from much of Africa, Asia, and Europe since the Late Oligocene. The earliest taxa, from Africa, resemble extant hominoids but share with them mainly primitive characters. Middle and Late Miocene taxa are clearly hominoids, and by the end of the Middle Miocene most can be attributed to either the pongine (Pongo) or hominine (African ape and human) clade. Interestingly, there is no fossil record of the hylobatid clade (gibbons and siamangs). Miocene hominoids experienced a series of dispersals between Africa, Europe, and Asia that mirror those experienced by many other contemporaneous land mammals. These intercontinental movements were made possible by the appearance of land bridges, changes in regional and global climatic conditions, and evolutionary innovations. Most of the attributes that define the hominids evolved in the expansive subtropical zone that was much of Eurasia. Hominines and pongines diverge from each other in Eurasia, and the final Miocene dispersal brings the hominine clade to Africa and the pongine clade to Southeast Asia. Having moved south with the retreating subtropics, hominines and pongines finally diverge in situ into their individual extant lineages.

4.1 Introduction

Nonhuman fossil hominoids represent a highly diverse and successful radiation of catarrhine primates known from many localities ranging geographically from Namibia in the South, Germany in the North, Spain in the West, and Thailand in the East, and temporally from Oligocene deposits in Kenya to the Pleistocene of China (Figure 4.1). More than 40 genera of nonhuman hominoids are known (Table 4.1), probably a small percentage of the total number that have existed. Given the focus of these volumes on ape and especially human evolution, this survey of the fossil record of Miocene hominoids will concentrate on taxa that most or all researchers agree are hominoid and in particular on taxa that are most informative on the pattern and biogeography of modern hominoid origins.
4.1.1 What is a hominoid?

Most of the fossil taxa attributed to the Hominoidea or the Hominidea (new rank, Table 4.2) in this chapter are known to share derived characters with living hominoids. Because the two living families of the Hominoidea, Hylobatidae and Hominidae, share characters that are either absent or ambiguous in their development in *Proconsul* and other Early Miocene taxa, a new rank is proposed here to express the monophyly of the Hominoidea and the monophyly of catarrhines more closely related to extant hominoids than to any other catarrhine. The magnafamily Hominidea (a rank proposed in a work on perissodactyl evolution [Schoch 1986]) unites Proconsuloidea with Hominidea to the exclusion of other catarrhines. This differs from Harrison’s use of the term proconsuloid that he sees as referring to the sister taxon to cercopithecoids and hominoids (Harrison 2002).
### Table 4.1

**Genera of fossil apes**

<table>
<thead>
<tr>
<th>Age</th>
<th>Ma</th>
<th>Genera</th>
<th>Important localities</th>
<th>Country</th>
<th>Material</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oligo.</td>
<td>25</td>
<td>Kamoyapithecus(^3)</td>
<td>Lothidok</td>
<td>Kenya</td>
<td>Craniodental fragments</td>
</tr>
<tr>
<td>e M</td>
<td>21</td>
<td>New taxon</td>
<td>Meswa Bridge</td>
<td>Kenya</td>
<td>Craniodental fragments</td>
</tr>
<tr>
<td>e M</td>
<td>?20–17.5</td>
<td>Morotopithecus</td>
<td>Moroto</td>
<td>Uganda</td>
<td>Cranial, dental, postcrania</td>
</tr>
<tr>
<td>e M</td>
<td>?20–17.5</td>
<td>Kogolepithecus(^3)</td>
<td>Moroto</td>
<td>Uganda</td>
<td>Dental</td>
</tr>
<tr>
<td>e M</td>
<td>19</td>
<td>Ugandaapithecus(^4)</td>
<td>Napan/Songhor</td>
<td>K/U</td>
<td>Cranial, dental, postcrania</td>
</tr>
<tr>
<td>e M</td>
<td>19</td>
<td>Xenopithecus(^3)</td>
<td>Koru</td>
<td>Kenya</td>
<td>Craniodental fragments</td>
</tr>
<tr>
<td>e M</td>
<td>19</td>
<td>Proconsul</td>
<td>Songhor/Koru</td>
<td>Kenya</td>
<td>Cranial, dental, postcrania</td>
</tr>
<tr>
<td>e M</td>
<td>19</td>
<td>Limnopithecus(^3)</td>
<td>Koru/Songhor</td>
<td>Kenya</td>
<td>Craniodental fragments</td>
</tr>
<tr>
<td>e M</td>
<td>19</td>
<td>Rangwapaniticus(^3)</td>
<td>Songhor</td>
<td>Kenya</td>
<td>Craniodental fragments</td>
</tr>
<tr>
<td>e M</td>
<td>19</td>
<td>Micropithecus(^3)</td>
<td>Napan/Koru</td>
<td>K/U</td>
<td>Craniodental fragments</td>
</tr>
<tr>
<td>e M</td>
<td>19</td>
<td>Kalepithecus(^3)</td>
<td>Songhor/Koru</td>
<td>Kenya</td>
<td>Craniodental fragments</td>
</tr>
<tr>
<td>e M</td>
<td>17.5–19</td>
<td>Dendropithecus(^3)</td>
<td>Rusinga/Songhor/Napan/Koru</td>
<td>K/U</td>
<td>Cranial, dental, postcrania</td>
</tr>
<tr>
<td>e M</td>
<td>17.5</td>
<td>cf. Proconsul(^6)</td>
<td>Rusinga/Mfangano</td>
<td>Kenya</td>
<td>(Cranial, dental, postcrania)+</td>
</tr>
<tr>
<td>e M</td>
<td>17.5</td>
<td>Turkanapithecus</td>
<td>Kalodirr</td>
<td>Kenya</td>
<td>Cranial, dental, postcrania</td>
</tr>
<tr>
<td>e M</td>
<td>17.5</td>
<td>Afropithecus</td>
<td>Kalodirr</td>
<td>Kenya</td>
<td>Cranial, dental, postcrania</td>
</tr>
<tr>
<td>e–m M</td>
<td>17.5–15</td>
<td>Simiulus</td>
<td>Kalodirr/Maboko</td>
<td>Kenya</td>
<td>Cranial, dental, postcrania</td>
</tr>
<tr>
<td>e–m M</td>
<td>17.5–15</td>
<td>Nyanzapithecus</td>
<td>Rusinga/Maboko</td>
<td>Kenya</td>
<td>Craniodental fragments</td>
</tr>
<tr>
<td>e M</td>
<td>17</td>
<td>Heliopithecus</td>
<td>Ad Dabtiyah</td>
<td>S. Arabia</td>
<td>Dental</td>
</tr>
<tr>
<td>e M</td>
<td>16.5</td>
<td>cf. Griphopithecus</td>
<td>Engelswies</td>
<td>Germany</td>
<td>Dental</td>
</tr>
<tr>
<td>e M</td>
<td>16</td>
<td>Griphopithecus</td>
<td>Paşalar/Çandır</td>
<td>Turkey</td>
<td>Cranial, (dental)+, postcrania</td>
</tr>
<tr>
<td>m M</td>
<td>15</td>
<td>Equatorius</td>
<td>Maboko/Kipsaramon</td>
<td>Kenya</td>
<td>(Cranial, dental, postcrania)+</td>
</tr>
<tr>
<td>m M</td>
<td>15</td>
<td>Mabokopithecus</td>
<td>Maboko</td>
<td>Kenya</td>
<td>Dental</td>
</tr>
<tr>
<td>m M</td>
<td>15</td>
<td>Nacholapithecus</td>
<td>Nachola</td>
<td>Kenya</td>
<td>Partial skeleton</td>
</tr>
<tr>
<td>m M</td>
<td>13</td>
<td>Pierolapithecus</td>
<td>El Hostelets de Pierola</td>
<td>Spain</td>
<td>Partial skeleton</td>
</tr>
<tr>
<td>m M</td>
<td>13</td>
<td>Kenyapithecus</td>
<td>Fort Ternan</td>
<td>Kenya</td>
<td>Cranial, dental, postcrania</td>
</tr>
<tr>
<td>m M</td>
<td>13</td>
<td>Otaviapithecus</td>
<td>Otavi</td>
<td>Namibia</td>
<td>Craniodental, vertebra</td>
</tr>
<tr>
<td>m–l M</td>
<td>12–7</td>
<td>Sivapithecus</td>
<td>Potwar Plateau</td>
<td>Pakistan</td>
<td>(Cranial, dental, postcrania)+</td>
</tr>
<tr>
<td>m–l M</td>
<td>12–10</td>
<td>Dryopithecus</td>
<td>Rudabánya/Can Llobateres</td>
<td>H/S</td>
<td>(Cranial, dental, postcrania)+</td>
</tr>
</tbody>
</table>
Table 4.1 (continued)

<table>
<thead>
<tr>
<th>Age</th>
<th>Ma</th>
<th>Genera</th>
<th>Important localities</th>
<th>Country</th>
<th>Material</th>
</tr>
</thead>
<tbody>
<tr>
<td>m–l M</td>
<td>13.5–7</td>
<td><em>Khoratpithecus</em>&lt;sup&gt;6&lt;/sup&gt;</td>
<td>Ban Sa/Khorat</td>
<td>Thailand</td>
<td>Craniodental fragments</td>
</tr>
<tr>
<td>I M</td>
<td>10</td>
<td><em>Ankarapithecus</em></td>
<td>Sinap</td>
<td>Turkey</td>
<td>Cranial, dental, postcrania</td>
</tr>
<tr>
<td>I M</td>
<td>9.5</td>
<td><em>Samburupithecus</em></td>
<td>Samburu</td>
<td>Kenya</td>
<td>Craniodental fragments</td>
</tr>
<tr>
<td>I M</td>
<td>9.5</td>
<td><em>Ouranopithecus</em></td>
<td>Ravin de la Pluie</td>
<td>Greece</td>
<td>(Craniodental)+, 2 phalanges</td>
</tr>
<tr>
<td>I M</td>
<td>9–8</td>
<td><em>Graecopithecus</em></td>
<td>Pygros</td>
<td>Greece</td>
<td>Mandible</td>
</tr>
<tr>
<td>I M</td>
<td>9–8</td>
<td><em>Lufengpithecus</em></td>
<td>Lufeng</td>
<td>China</td>
<td>(Cranial, dental)+, postcrania</td>
</tr>
<tr>
<td>I M</td>
<td>8–7</td>
<td>New taxon</td>
<td>Çorakyerler</td>
<td>Turkey</td>
<td>Mandible, maxilla</td>
</tr>
<tr>
<td>I M</td>
<td>7</td>
<td><em>Oreopithecus</em></td>
<td>Baccinello/Monte Bamboli</td>
<td>Italy</td>
<td>(Cranial, dental, postcrania)+</td>
</tr>
<tr>
<td>I M</td>
<td>7–6</td>
<td><em>Sahelanthropus</em></td>
<td>Toros-Menalla</td>
<td>Chad</td>
<td>Craniodental</td>
</tr>
<tr>
<td>I M</td>
<td>6.5</td>
<td><em>Gigantopithecus</em></td>
<td>Potwar Plateau</td>
<td>Pakistan</td>
<td>Mandible</td>
</tr>
<tr>
<td>I M</td>
<td>6</td>
<td><em>Orrorin</em></td>
<td>Lukeino</td>
<td>Kenya</td>
<td>Craniodental, postcrania</td>
</tr>
<tr>
<td>I M</td>
<td>5.8–5.2</td>
<td><em>Ardipithecus</em>&lt;sup&gt;7&lt;/sup&gt;</td>
<td>Alayla (Middle Awash)</td>
<td>Ethiopia</td>
<td>Craniodental, postcrania</td>
</tr>
</tbody>
</table>

Oligo. = Oligocene, e M = Early Miocene, e–m M = Early–Middle Miocene, m M = Middle Miocene, m–l M = Middle–Late Miocene, I M = Late Miocene, K/U = Kenya and Uganda, H/S = Hungary and Spain.

<sup>1</sup>These 40 genera include taxa from the Oligocene and Early Miocene that share mainly primitive characters with the Hominidea but that appear to be derived relative to Pliopithecoids and Propliopithecoids.

<sup>2</sup>Material briefly described by part representation. Dental = mainly isolated teeth; Dental, mandible, maxilla = known only from these parts; Craniodental fragments = teeth and few cranial fragments; Craniodental = larger samples of more informative cranial material; Cranial, dental, postcrania = good samples from each region; ( )+ = very good representation of parts in parentheses.

<sup>3</sup>Unclear attribution to Hominidea.

<sup>4</sup>Large taxon possibly distinct from *Proconsul*.

<sup>5</sup>*Proconsul* from later Early Miocene localities in Kenya is likely to be distinct from earlier Miocene *Proconsul* type material.

<sup>6</sup>Middle Miocene samples attributed to this taxon are more fragmentary and may not be congeneric.

<sup>7</sup>*Ardipithecus kadabba*. 
A few taxa are included in this review if they are too poorly known to preserve unambiguous hominoid synapomorphies but closely resemble other better-known fossil hominoids. In general, fossil and living hominoids retain a primitive catarrhine dental morphology. This makes it difficult to assign many fossil taxa to the Hominidea since a large number are known only from teeth and small portions of jaws. Dentally, the most primitive Hominidea differ only subtly from extinct primitive catarrhines (propliopithecoids and pliopithecoids) (Figure 4.2, node 1). Propliopithecoids (Propliopithecus, Aegyptopithecus) are usually smaller and have much more strongly developed molar cingula, higher cusped premolars, and smaller incisors and canines (Begun et al. 1997; Rasmussen 2002). Pliopithecoids (Pliopithecus, Anapidicus) are also generally smaller and have molars with more strongly expressed cingula, more mesial protoconids,
and relatively small anterior teeth (Begun 2002). However, the differences between Late Miocene hominids and Late Miocene pliopithecoids are more marked than between Early Miocene Hominidea and pliopithecoids, making defining features less than clear-cut. Cranially Hominidea have a completely ossified tubular ectotympanic, which distinguishes them from both propliopithecoids and pliopithecoids but not from cercopithecoids. However, few Miocene Hominidea fossils preserve this region. Hylobatids, cranially the most primitive extant hominoid, share many features found in short-faced Old and New World monkeys, again making it difficult to tease out synapomorphies. Hominoids show a tendency to expand the length and superoinferior thickness or robusticity of the premaxilla, with increasing overlap with the palatine process of the maxilla over time, but once again this is not present in hylobatids or early well-preserved specimens of Proconsul, for example (Begun 1994a). Only one specimen of Early Miocene Hominidea is complete enough to say much about the brain, and there are no unambiguous synapomorphies linking it to hominoids. The brain of Proconsul is similar relative to body size to both hylobatids and papionins, the Old World monkeys with the largest brains, and the sulcal pattern, while debatable, lacks most if not all hominoid features (Falk 1983; Begun and Kordos 2004). Like hylobatids, most cercopithecoids, and most mammals other than hominids, a portion of the brain of Proconsul occupied a large subarcuate fossa. Cranial and dental evidence also suggests that Proconsul was moderately delayed in terms of life history, another similarity with extant hominoids (Kelley 1997, 2004).

Postcranially, Proconsul more clearly represents the ancestral hominoid morphotype, though this too is the subject of debate. Proconsul fossils exhibit hominoid attributes of the elbow, wrist, vertebral column, hip joint, and foot, though in all cases these are subtle and disputed (Beard et al. 1986; Rose 1983, 1988, 1992, 1994, 1997; Ward et al. 1991; Ward 1993, 1997; Begun et al. 1994; but see Harrison 2002, 1987). Proconsul has a suite of characters consistent with the hypothetical ancestral morphotype of the hominoids, and it should not be surprising that these are poorly developed at first, only to become more refined as hominoids evolve. In comparison to the hominoid outgroup (cercopithecoids), we can expect the earliest hominoids to show subtle indications of increased orthogrady, positional behaviors with increased limb flexibility and enhanced grasping capabilities and no tail, generalized (primitive) dentition, encephalization at the high end of extant cercopithecoids of comparable body mass, and life history variables closer to extant hominoids than to extant cercopithecoids. Proconsul has all of these attributes.

If these are the features that define the Hominidea, which taxa among Miocene fossil catarrhines are not Hominidea? Even the earliest cercopithecoids
(victoriapithecids) are easily distinguished from hominoids (Benefit and McCrossin 2002). Pliopithecoids, often grouped with the “apes,” are even more distantly related. They are clearly stem catarrhines lacking synapomorphies of all crown catarrhines including Proconsul and Victoriapithecus (Begun 2002). The most informative among these synapomorphies are the tubular ectotympanic and the entepicondylar groove (often referred to as the absence of an entepicondylar foramen). In the following sections, I will summarize current knowledge of the Miocene Hominidea, focusing on well-known taxa that serve to illustrate important events in hominoid evolutionary history (Figure 4.2).

4.2 Origins of Hominidea

It is likely that hominoids originated in Africa from an ancestor that, if known, would be grouped among the Pliopithecoida. Pliopithecoids, currently known only from Eurasia, share with all catarrhines the same dental formula and possibly with crown catarrhines a reduction of the midface, subtle features of the molar dentition, and a partial ossification of the ectotympanic tube (Begun 2002). The presence of pliopithecoids in Africa is suggestive but remains to be demonstrated (Andrews 1978; Begun 2002). The oldest and most primitive catarrhine that can lay claim to hominoid status however is African (Table 4.1). Kamoyapithecus, from the Oligocene of Kenya, differs from other Oligocene catarrhines (propliopithecoids) in being larger and having canines and premolars that more closely resemble Miocene hominoids than Oligo-Miocene non-hominoids (Leakey et al. 1995). Only craniodental material of Kamoyapithecus has been described, and it is so primitive as to make attribution to the Hominoidea difficult. Though it would fail to fall among the Hominoidea in a quantitative cladistic analysis due to its fragmentary preservation and primitive morphology, it makes in my view a good hominoid precursor.

4.3 Proconsuloidea

4.3.1 Proconsul

The superfamily Proconsuloidea, as defined by Harrison (2002), includes many mainly Early Miocene taxa. As noted, in this chapter a number of taxa from this group are interpreted to represent primitive Hominidea or hominoids. A hypothetical ancestral morphotype for the Hominidea is given in Figure 4.2 (node 1). Node 1 represents the bifurcation of Proconsul from hominoids with more
apparent synapomorphies to living hominoids. *Proconsul* as described here is based mainly on the sample from Rusinga Island, Kenya along with other localities of the same age. These include the species *Proconsul heseloni* (Walker et al. 1993) and *Proconsul nyanzae* (Le Gros Clark and Leakey 1950). The type
specimen of *Proconsul africanus* (Hopwood 1933) is from Koru, which is 1-to 2-Myr older, as is the type of *Proconsul major* (Le Gros Clark and Leakey 1950) from Songhor (Drake et al. 1988). There is evidence that the younger species of *Proconsul* that are the basis of the description here belong to a different genus from the older type sample (see below). However, as this is not the appropriate venue to name a new genus, I will follow convention and refer to the Rusinga sample as *Proconsul*.

### 4.3.1.1 Postcranial morphology

*Proconsul* and other proconsuloids are defined by a large number of characters. *Proconsul* is a generalized arboreal quadruped but is neither monkey-like nor apelike (Rose 1983). The following summary is mainly from Rose (1997), Ward (1997), and Walker (1997). In addition to the characters noted that emphasized its hominoid affinities, *Proconsul* has limbs of nearly equal length, with scapula positioned laterally on the thorax and the ovoid and narrow glenoid positioned inferiorly, as in generalized quadrupeds. The thorax is transversely narrow and deep superoinferiorly, and the vertebral column is long and flexible, especially in the lumbar region. The innominate is long with a narrow ilium and an elongated ischium. The sacrum is narrow, and its distal end indicates that it articulated with a coccygeal and not a caudal vertebra, in other words *Proconsul* had a coccyx and not a tail (Ward et al. 1991).

In the details of limb morphology, *Proconsul* also combines aspects of monkey and ape morphology. *Proconsul* forelimbs lack the characteristic elongation of ape forelimbs. The humeral head is oriented posteriorly relative to the transverse plane and the humeral shaft is convex anteriorly, both of which are consistent with the position of the glenoid fossa and the shape of the thorax. The distal end of the humerus lacks the enlargement of the capitulum and trochlea and other details of the hominoid elbow, but it does have a narrow zona conoidea and a mild trochlear notch. The medial epicondyle is also more posteriorly oriented as in monkeys. The proximal ends of the radius and ulna are consistent with the morphology of the distal humerus. The radial head is small and ovoid, the ulnar trochlea is narrow and has a poorly developed keel, and the radial notch is positioned anteriorly. The ulna also has a large olecranon process. All of these features are consistent with generalized pronograde (above branch) quadrupedalism as opposed to antipronograde (suspensory or below branch) (see Ward Volume 2 Chapter 6).

Distally, the radial carpal surface is flat and articulates mainly with the scaphoid. The ulnar head is comparatively large with a long and prominent
styloid process that articulates directly with the pisiform and triquetrum, unlike living hominoids, which have greatly reduced ulnar styloids and no contact with the carpals. The carpals are small transversely. The scaphoid is separate from the os centrale and the midcarpal joint is narrow. The hamate hamulus is small, and the surface for the triquetrum is flat and mainly medially oriented (Beard et al. 1986). The metacarpal surfaces of the distal carpals are small and comparatively simple as are the corresponding surfaces on the metacarpal bases. The metacarpals are short and straight and their heads transversely narrow. The proximal ends of the proximal phalanges are slightly dorsally positioned as in palmigrade quadrupeds. All the phalanges are short and straight compared to apes, though secondary shaft features, in particular of the proximal phalanges, suggest powerful grasping (Begun et al. 1994).

The hindlimbs of *Proconsul* are also dominated by monkey-like characters. The long bones are long and slender. The femoral head is small compared to apes, but its articulation with the acetabulum indicates more mobility compared to most monkeys. The feet of *Proconsul* are monkey-like in their length to breadth ratio (they are narrow compared to great ape feet). *Proconsul* tarsals are elongated relative to breadth and the metatarsals long compared to the phalanges. Like those of the hands, the foot phalanges of *Proconsul* are straighter and less curved than in apes but with more strongly developed features related to grasping than in most monkeys. The hallucial phalanges are relatively robust, suggestive of a powerfully grasping big toe. Body mass estimates for the species of *Proconsul*, based mainly on postcranial evidence, range from about 10 to 50 kg (Ruff et al. 1989; Rafferty et al. 1995).

### 4.3.1.2 Craniodental morphology

As noted, *Proconsul* has a moderate amount of encephalization (comparable to hylobatids and papionins), a short face with a fenestrated palate (Figure 4.3), a smoothly rounded and somewhat airorhynchous face (Figure 4.4), and a generalized dentition. Morphologically, the dentition is consistent with a soft fruit diet, and microwear analysis suggests the same (Kay and Ungar 1997). The somewhat enlarged brain of *Proconsul* implies a degree of life history delay approaching the hominoid pattern (Kelley 1997, 2004).

One aspect of the cranium of *Proconsul* that has received some attention is the frontal sinus. Walker (1997) interprets the presence of a frontal sinus in *Proconsul* to indicate its hominid status, citing the presence of large frontal sinuses in some great apes. Other researchers have suggested that the frontal
sinus is a primitive character, as it is found in *Aegyptopithecus* and many New World monkeys (Andrews 1992; Rossie et al. 2002). The confusion stems from the use of one term to describe several different characters. As Cave and Haines (1940) noted long ago, frontal sinuses in primates have various ontogenetic origins, and it is likely that they are not homologous across the primates. New World monkeys have frontal sinuses that are outgrowths from the sphenoid sinus, as is also the case for hylobatids. On the other hand, *Pongo*, which occasionally has a frontal sinus, derives it from the maxillary sinus. African apes and humans normally have large frontal sinuses derived from the ethmoidal sinuses. “Frontal sinuses” then are actually three different characters, frontosphenoidal sinuses, frontomaxillary sinuses, and frontoethmoidal sinuses.

While it is possible to establish the ontogenetic origin of a pneumatized frontal bone in living primates, it is more difficult in fossil primates. However, the placement and size of the frontal sinuses correlate very well with their ontogenetic

**Figure 4.3**

Midsagittal cross section of a number of Hominidea palates showing some of the features described in the text. *Proconsul* has a small premaxilla and a fenestrated palate (large foramen and no overlap between the maxilla and premaxilla). *Nacholapithecus* has a longer premaxilla with some overlap. It is similar to *Afropithecus* and conceivably could be the primitive morphotype for the Hominidae. *Pongo* and *Sivapithecus* have a similar configuration but with further elongation and extensive overlap between the maxilla and premaxilla, producing a smooth subnasal floor. Hominines have robust premaxillae that are generally shorter and less overlapping than in *Sivapithecus* and *Pongo*. *Dryopithecus* is most similar to *Gorilla*, which may represent the primitive condition for hominines. *Pan* and *Australopithecus* have further elongation and overlap, but the configuration differs from *Pongo*. This morphology is suggested to be an important synapomorphy of the *Pan/Homo* clade (Begun 1992b). Modified from Begun (1994)
origin, offering a protocol for identifying the specific type of frontal sinus present in a fossil (Begun 1994a). Frontosphenoidal sinuses invade large portions of the frontal squama but not the supraorbital or interorbital regions. Frontomaxillary sinuses are infrequent in *Pongo*, but when they occur they are associated with narrow canals or invaginations connecting the maxillary sinuses to a small pneumatization of the frontal via the interorbital space. In African apes and humans, the frontoethmoidal sinuses arise from a spreading of the ethmoidal air cells in the vicinity of nasion, resulting in large pneumatizations from below nasion into the supraorbital portion of the frontal. The actual amount of frontal pneumatization is variable, while the presence of a large sinus around nasion is constant. Therefore, while we cannot observe the development of frontal pneumatization in fossil primates, and we do not have adequate ontogenetic series to directly reconstruct this growth, we can infer the type of frontal pneumatization from its position, extent, and connection to the source sinus. In *Proconsul*, as in hyllobatids, New World monkeys, and *Aegyptopithecus*, the frontal pneumatization is
extensive and occupies the frontal squama, consistent with a frontosphenoidal sinus. Thus, the “frontal sinus” in Proconsul is a primitive character, as suggested by Andrews (1992), but for different reasons. The frontal pneumatization of Dryopithecus, on the other hand, conforms to the pattern seen exclusively in African apes and humans (see below).

In summary, Proconsul was an above branch mid- to large-sized catarrhine with a diet dominated by soft fruits and a somewhat slower life history than cercopithecoids. Encephalization may imply other similarities to hominoid behavioral or social ecology, or it may simply be a consequence of relatively large body mass and/or a slower life history (Kelley 2004; Russon and Begun 2004). The slightly enhanced range of motion in Proconsul limbs may imply some degree of orthogrady, or it may be a consequence of lacking a tail or of large body mass in an arboreal milieu, or some combination of all three (Beard et al. 1986; Begun et al. 1994; Kelley 1997).

4.3.2 Other possible proconsuloids

A number of taxa are regarded by many researchers as having a probable close relationship to Proconsul. The three with the best evidence for affinities to the proconsuloids are Proconsul sensu stricto, Micropithecus, and Samburupithecus. As noted, the type species of the genus Proconsul is P. africanus. P. africanus and P. major, which are both older than Proconsul from Rusinga, are also more primitive and lack synapomorphies shared by Proconsul from Rusinga and other hominoids (see below). Other taxa listed in Table 4.2 are either more likely to be hominoids given similarities to known hominoids (Rangwapithecus, Nyanzapithecus, Mabokopithecus) or they are so primitive or poorly known as to cast doubt on their magnafamily status.

4.3.2.1 Proconsul sensu stricto

The older species of Proconsul sensu stricto from Songhor and Koru (P. africanus and P. major) probably represent a different genus from P. heseloni and P. nyanzae, the samples on which the descriptions of Proconsul presented here are based. A new genus would replace P. heseloni and P. nyanzae, as P. africanus has priority. Proconsul sensu stricto from Songhor and Koru has elongated postcanine teeth, more strongly developed cingula, upper premolars with strong cusp heteromorphy and conical, individualized molar cusps, all of which suggest that the older species are in fact more primitive. The two older species differ from each other
only in size. Postcrania attributed to Proconsul from the older Proconsul sensu stricto localities are distinct from postcrania from the younger Proconsul localities such as Rusinga, in details that have been correlated to paleoecological differences (Andrews et al. 1997). While sufficiently similar to the better-known younger Proconsul sample to warrant placing both in the same superfamily, the more modern morphology of the younger Proconsul sample will almost certainly require taxonomic recognition.

Ugandapithecus, a recently named genus based on the sample of P. major, adds to the confusion (Senut et al. 2000). P. africanus is smaller than “Ugandapithecus” major but is morphologically closer to it in the attributes noted above. Thus, while Ugandapithecus is probably a junior subjective synonym of Proconsul sensu stricto, it turns out that Proconsul as traditionally defined probably does represent more than one genus. Senut et al. (2000) have suggested that large-bodied hominoids from Moroto, in Uganda, may also be attributed to Ugandapithecus (hence the name, though the type is from Kenya), that is, P. major, calling into question the interpretation that hominoid cranial and postcranial fossils from Moroto belong to one taxon (Morotopithecus). However, the evidence for more than one large hominoid genus at Moroto is not strong (see below).

4.3.2.2 Micropithecus

Micropithecus (Fleagle and Simons 1978) is a small catarrhine with comparatively broad incisors and long postcanine teeth with low cusps and rounded occlusal crests. The cingula are less strongly developed than most other Early Miocene catarrhines. Males and females exhibit marked size dimorphism. Comparisons to living catarrhines suggest a body mass of about 34.5 kg (Harrison 2002). While Harrison (2002) considers this taxon to be even more distantly related to the Hominoidea than are the Proconsuloida, the subtly more modern features of the dentition suggest that it may belong to the Proconsuloida. Fleagle and Simons (1978) in fact attribute Micropithecus to the Hominoidea, although as noted the features shared with hominoids are very subtle. If Micropithecus is a proconsuloid, as suggested here, it would indicate that the proconsuloids were quite diverse in body mass, as is the case in all catarrhine superfamilies.

4.3.2.3 Samburupithecus

Samburupithecus is another possible proconsuloid, known only from a large maxillary fragment from the Late Miocene of the Samburu region of Kenya
(Ishida and Pickford 1997). Ishida and Pickford (1997) have suggested that Samburupithecus is an early member of the African ape and human clade. However, Samburupithecus retains many primitive characters of the Proconsulidea. These include a low root of the zygomatic processes, a strongly inclined nasal aperture edge, the retention of molar cingula, and thick enamel with high dentine relief (Begun 2001). Samburupithecus is most likely to be a late surviving proconsuloid. Its unusual dental characters (e.g., large molars with individualized cusps separated by deep narrow fissures) are reminiscent of morphological “extremes” found in terminal lineages with long evolutionary histories. Oropithecus, Gigantopithecus, Paranthropus, Daubentonia, and Ekgmowechashala all share with Samburupithecus exaggerated occlusal features compared to other members of their respective clades. The size and occlusal morphology of Samburupithecus that superficially resembles Gorilla (Ishida and Pickford 1997) may be related in part to the fact that both are ends of long phylogenetic branches. In a pattern analogous to long branch attraction in molecular systematics, there is a tendency for separate long isolated lineages to converge in certain aspects of their morphology (Begun 2001). For whatever the reason, in its details Samburupithecus is primitive and more likely to belong to the proconsuloids than the hominoids.

4.4 Early hominoids

4.4.1 Afropithecus

A number of late Early Miocene and Middle Miocene taxa share characters with extant Hominoids and are included here in the Hominoidea. Afropithecus (Leakey and Leakey 1986) is known from several localities in northern Kenya dated to 17–17.5 Ma (Leakey and Walker 1997). Afropithecus shares an increase in premaxillary robusticity and length with most extant hominoids (Figure 4.3). Like many Late Miocene and Pliocene hominids, Afropithecus has very thick occlusal enamel as well. On the other hand, Afropithecus cranial morphology, particularly the morphology of the midface, recalls that of Aegyptopithecus. Thus, as with Proconsul, Afropithecus has a mosaic of primitive and derived characters (Leakey et al. 1991; Leakey and Walker 1997). Leakey and Walker (1997) have suggested that the unusual primitive looking face of Afropithecus may be related to a specialized scerocarp seed predator adaptation. This is functionally consistent with the robust, prognathic premaxilla, large, relatively horizontal incisors, large but relatively low-crowned canines, expanded premolars, thick enamel, and powerful chewing muscles of Afropithecus. A similar set of features
is found in modern primate seed predators such as pithecines. It may be that some of the primitive appearance of the *Afropithecus* face is homoplastic with *Aegyptopithecus*.

A large cranial specimen of *Afropithecus* reveals some information on neurocranial morphology. KNM-WK 16999 preserves a portion of the braincase immediately behind the orbits. Postorbital constriction is very marked as are the anterior temporal lines. These converge to form a pronounced and very anteriorly situated sagittal crest. The small portion of the anterior cranial fossa preserved indicates a cerebrum that was constricted rostrally, lacking the frontal lobe expansion typical of extant and fossil great apes.

*Afropithecus* is similar in size to *Proconsul nyanzae*, based on cranial, dental, and postcranial dimensions (Leakey and Walker 1997). *Afropithecus* postcrania are very similar to those of *Proconsul*, so much so that Rose (1993) found them essentially indistinguishable. The postcranial adaptation, in terms of body mass and positional behavior, is *Proconsul*-like, while the craniodental anatomy is markedly distinct.

Another early hominoid taxon, *Heliopithecus*, is contemporaneous with *Afropithecus* and morphologically very similar, though it is only known from a fragmentary hemimaxilla and a few isolated teeth (Andrews and Martin 1987). Both *Heliopithecus* and *Afropithecus* share characters that are found next in the fossil record in Eurasia, which suggests that *Heliopithecus* and *Afropithecus* taxa may have a closer relationship to late Early Miocene and Middle Miocene hominoids from Europe than do proconsuloids.

### 4.4.2 *Morotopithecus*

*Morotopithecus* is a fossil hominoid from Uganda dated to about 21 Ma by some and 17 Ma by others (Gebo et al. 1997; Pickford et al. 2003). It is best known from a large cranial specimen including most of the palate (Pilbeam 1969). For many years, this specimen was attributed to *Proconsul major*, but it is clear that it and other specimens from Moroto are sufficiently distinct to justify a new genus. *Morotopithecus* is similar in size to *Proconsul major* and larger than *Proconsul nyanzae* and *Afropithecus*. It lacks the distinctive subnasal morphology of *Afropithecus* and has a *Proconsul*-like premaxilla that is short, gracile, and does not overlap the palatine process of the maxilla, resulting in a large incisive fenestration (Figure 4.3). *Morotopithecus* has a broad palate, large anterior teeth, especially the canines, which are tusklike, a piriform aperture broadest about mid way up, and an interorbital space that appears to be relatively narrow, though it is damaged. However, the most important distinction of *Morotopithecus* is the
morphology of the postcrania, which are said to be modern hominoid-like. Newly described specimens of *Morotopithecus* include the shoulder joint, hip joint, and details of the vertebral column. Walker and Rose (1968) described the vertebrae as hominoid-like, which has been confirmed by more recent discoveries and analyses. The glenoid fossa of *Morotopithecus* suggests a more mobile shoulder joint as does the morphology of the hip joint. However, it is the hominoid-like position of the transverse processes of the vertebrae that represents the strongest evidence for the hominid affinities of *Morotopithecus*. The roots of the transverse processes of the lumbar vertebrae of *Morotopithecus* are positioned posteriorly, as in extant great apes, suggesting a stiff lower back and an axial skeleton like that of extant hominoids.

### 4.4.3 Eurasian hominoid origins

As noted, *Heliopithecus* and *Afropithecus* have more robust jaws and teeth than Early Miocene proconsuloids, and this may represent a key adaptation that permitted the expansion of hominoids into Eurasia at about 17 Ma, when during a marine low stand a diversity of terrestrial mammals moved between Africa and Eurasia (Made 1999; Begun 2001; Begun et al. 2003). Toward the end of the Early Miocene, the movement of the southern landmasses northward, combined with a number of other developments (the Alpine and Himalayan orogenies, the earliest appearance the polar ice caps, and the Asian Monsoons), leads to a sequence of connections and barriers to terrestrial faunal exchange (Rögl 1999a, b; Adams et al. 1999; MacLeod 1999; Hoorn et al. 2000; Zhisheng et al. 2001; Guo et al. 2002; Liu and Yin 2002; Wilson et al. 2002). This period of global turbulence affected sea levels between continents cyclically such that for the remainder of the Miocene there would be periodic connections (low stands) and disconnections (high stands) between the continents. At about 17 Ma, a low stand that had permitted the exchange of terrestrial faunas between Eurasia and Africa (the Proboscidean datum) was coming to an end but not before hominoids possibly resembling *Afropithecus* and *Heliopithecus* dispersed from Africa into Eurasia (Heizmann and Begun 2001; Begun 2002, 2004; Begun et al. 2003).

At the end of the Early Miocene, about 16.5 Ma, hominoids of more modern dental aspect first appear in Eurasia. The oldest Eurasian hominoid is cf. *Griphopithecus*, known from a molar fragment from Germany. *Griphopithecus* (Abel 1902) is known mainly from large samples from Turkey of roughly the same age, while the type material is known from a probably later (14–15 Ma) locality, Děvinská Nová Ves, in Slovakia (Heizmann 1992; Andrews et al. 1996; Heizmann and Begun 2001; Begun et al. 2003). cf. *Griphopithecus* from Engelsweis in
Germany is a tooth fragment that has the more modern features of being thickly enameled with low dentine penetrance (tall dentine horns did not project into the thick enamel cap as in *Proconsul* and probably *Afropithecus*). cf. is the designation for a taxon that is similar enough to another taxon for there to be a strong likelihood that they are the same, but with some formally acknowledged uncertainty. *Griphopithecus* is better known from over 1,000 specimens, mostly isolated teeth, from 2 localities in Turkey (Çandır and Paşalar). One species, *Griphopithecus alpani* (Tekkaya 1974) is known from both localities, while a second somewhat more derived taxon is also found at Pasalar (Alpagut et al. 1990; Martin and Andrews 1993; Kelley and Alpagut 1999; Ward et al. 1999; Kelley et al. 2000; Kelley 2002; Güleç and Begun 2003).

*Griphopithecus alpani* has robust mandibles with strongly reinforced symphyses, broad, flat molars with thick enamel, and reduced cingulum development compared to *Proconsul*. It retains primitive tooth proportions (small M1 relative to M2), anterior tooth morphology, and postcanine occlusal outline shape. A few fragments of the maxilla from Pasalar indicate a primitive morphology for the anterior palate (Martin and Andrews 1993). The morphology of *Griphopithecus* molars as well as their microwear indicates a diet allowing the exploitation of hard or tough fruits, though it is not clear if this means simply that they could exploit these resources when needed (a keystone resource) or if they were a favored source of food. Two postcranial specimens, a humeral shaft and most of an ulna, from the younger site of Klein Hadersdorf, Austria, are also similar to *Proconsul* and indicate that *Griphopithecus* was a large bodied above branch arboreal quadruped similar in size and positional behavior to *Proconsul nyanzae* (Begun 1992a).

4.4.4 East African Middle Miocene thick-enamedled hominoids

Shortly after the appearance of *Griphopithecus* in Western Eurasia, dispersals between Eurasia and Africa were interrupted by the Langhian transgression (Rögl 1999a). Following the Langhian, at about 15 Ma, dispersals resumed in a number of mammal lineages, probably also including hominoids (Begun et al. 2003a, b). Hominoids closely similar to *Griphopithecus* in dental morphology appear in Kenya at this time. *Equatorius* is known from 15-Ma localities in the Tugen Hills and at Maboko, both in Kenya. This taxon, previously attributed to *Kenyapithecus*, is very similar to *Griphopithecus* but has a distinctive incisor morphology and reduced cingula, probably warranting a distinct genus status (Ward et al. 1999; Kelley et al. 2000; Ward and Duren 2002; contra Begun 2000,
Equatorius is also known from a good sample of postcrania, including most of the bones of the forelimb, vertebral column, hindlimb long bones, and a few pedal elements (Ward and Duren 2002). Like Grippopithecus, which is much less well known postcranially, Equatorius is similar to Proconsul nyanzae in postcranial size and morphology.

Nacholapithecus (Ishida et al. 1999) is known from deposits of the same age in the Samburu Basin of Kenya (Nakatsukasa et al. 1998). Like Equatorius, Nacholapithecus is known from a relatively complete skeleton, more complete in fact than the best specimen of Equatorius. From this exceptional skeleton, we know that Equatorius is similar to other Middle Miocene hominoids in most aspects of the jaws and teeth but unique in aspects of limb morphology. Kunimatsu et al. (2004) provide evidence that the anterior portion of the palate of Nacholapithecus is hominid-like in its length and degree of overlap with the palatine process of the maxilla (Figure 4.3). This morphology is the principal evidence for the hominoid status of this species. However in its details, the anterior palate of Nacholapithecus is unlike that of hominids (see below).

The postcranial skeleton of Nacholapithecus is similar to other Middle Miocene hominoids in having the general signature of a generalized, palmigrade, arboreal quadruped, but differs from Equatorius, also known from fore and hindlimb, in the enlarged size and robusticity of its forelimb. While not like extant nonhuman hominoids in forelimb length relative to the hindlimb, Nacholapithecus forelimbs are large and powerful, indicating a form of positional behavior emphasizing powerful forelimb grasping (Ishida et al. 2004).

Kenyapithecus (Leakey 1962) is the most derived of the Middle Miocene African hominoids and may be the earliest hominid (Table 4.2). Kenyapithecus is known only from a small sample from Fort Ternan in Kenya, though a second species may be present in Turkey (see below). Like other Middle Miocene hominoids, Kenyapithecus has large flat molars with broad cusps and thick enamel. The maxilla of Kenyapithecus, however, is derived in having a high root of the zygomatic, a probable hominid synapomorphy. While McCrossin and Benefit (1997) believe that Equatorius and Kenyapithecus represent a single species, most researchers have concluded that two genera are present and that Kenyapithecus is derived relative to Equatorius (Harrison 1992; Ward et al. 1999), and that is the view adopted here. As noted, it has been suggested that Kenyapithecus is also present at Pasalar. If so, and if Kenyapithecus is indeed an early hominid, this would date the origin of the hominid family to at least 16 Ma. However, there is a roughly 3 Ma gap between possible Kenyapithecus at Paşalar and the type material from Fort Ternan.

One last Middle Miocene hominoid deserves mention here. Otavipithecus is the only Miocene hominoid known from southern Africa, from the 13-Ma site of
Berg Aukas in Namibia (Conroy et al. 1992). Several specimens have been described, including the type mandible, a frontal fragment, and a few postcrania (Conroy et al. 1992; Pickford et al. 1997; Senut and Gommery 1997). It has been suggested that *Otavipithecus* has affinities to hominids (Conroy et al. 1992; Ward and Duren 2002), but it preserves primitive proconsuloid characters such as a small M₁ compared to M₂, a long M₃, parallel tooth rows, a small space for the mandibular incisors, low P₄ talonid, and tall, centralized molar cusps (Begun 1994b). Singleton (2000) carried out the most comprehensive analysis of *Otavipithecus* and concluded that it may be related to *Afropithecus*, which is broadly consistent with the placement of the taxon in Figure 4.2.

### 4.4.5 Summary of Middle Miocene hominoid evolution

The radiation of Middle Miocene hominoids in Africa was short lived. Having apparently dispersed from Eurasia by about 15 Ma, they are extinct by 12.5–13 Ma. Aside from a few fragmentary specimens that most likely represent the end of the *Proconsul* lineage (Hill and Ward 1988), hominoids would not appear again in Africa until the latest Miocene when they are represented by the earliest known hominins. Given the rarity of hominoid localities in the early Middle Miocene, the biogeographic hypothesis of the dispersal of Middle Miocene hominoids presented here is debatable. It is certainly possible, for example, that Early Miocene *Griphopithecus*-like fossils will be found in Africa that will show that *Equatorius, Nacholapithecus*, and *Kenyapithecus* all evolved *in situ* in Africa and that apparently earlier fossils from Eurasia are either misdated or are early offshoots of this clade with no direct relationship to later hominids (Ward and Duren 2002). However, the evidence as it currently stands supports the scenario presented here to be tested by new discoveries.

### 4.5 Early hominids

While *Kenyapithecus* shares a synapomorphy with the Hominidae (position of the zygomatic root), the first clear-cut hominids are known from Eurasia and share numerous cranial, dental, and postcranial synapomorphies with living hominids. The extant Hominidae is divided into two subfamilies, Ponginae and Homininae, and both are represented by roughly contemporaneous fossil taxa. The earliest hominines are represented by *Pierolapithecus, Dryopithecus*, and *Ouranopithecus*, all with affinities to fossil hominins and extant hominines, and the earliest pongines are represented by *Ankarapithecus, Sivapithecus*, and relatives. Miocene
hominines are known from western Eurasia, while Miocene pongines are known from South and East Asia, reflecting the basic biogeographic division of the two hominid subfamilies today.

### 4.5.1 Fossil pongines

The oldest sample of fossils widely interpreted as pongine is *Sivapithecus* from the middle Chinji formation of Siwaliks of Pakistan (Raza et al. 1983; Rose 1984, 1989; Kappelman et al. 1991; Barry et al. 2002). Specimens referred *Sivapithecus indicus* that are known to come from the middle Chinji formation share characters with later *Sivapithecus* and other hominids including reduced or absent molar cingula, relatively large M1, reduced premolar cusp heteromorphy, long, buccolingually compressed canines, broad based nasal aperture, elongated and robust premaxilla partly overlapping the maxilla and, as in *Kenyapithecus*, a high position to the zygomatic root. They share specifically with later *Sivapithecus* fewer clear-cut characters, such as probably strongly heteromorphic upper incisors (known only from the roots), and broad, flat cusped molars with thick enamel, though these characters are also found in most other Middle and many Late Miocene hominoids. One specimen of Chinji *Sivapithecus*, GSP 16075, represents a portion of the palate with the connection between the maxilla and premaxilla partially preserved. The maxillary–premaxillary relationship is highly diagnostic of *Sivapithecus* and the pongine clade, and the morphology of the Chinji specimen has been interpreted to share characters of this complex (Raza et al. 1983; Ward 1997; Kelley 2002). However, while the specimen does have a relatively elongated, horizontal, and robust premaxilla, the area of the incisive fossa and foramen are not preserved. In the absence of this region, it is difficult to distinguish Chinji *Sivapithecus* from later *Sivapithecus*, including later *Sivapithecus indicus*, to which it is assigned, versus another pongine, *Ankarapithecus*. Resolution of this uncertainty may help to clarify the biogeography of pongine origins (see below).

### 4.5.2 *Sivapithecus*

#### 4.5.2.1 *Sivapithecus* craniodental evidence

Most of the *Sivapithecus* samples, including the best-known specimens with the clearest evidence of pongine affinities, are from younger localities of the Siwaliks of India and Pakistan, dated between 10.5 and 7.5 Ma (Barry et al. 2002). In the
following section, I discuss *Sivapithecus* in some detail because it is in many ways critical to understanding Late Miocene hominoid evolution. Three species are generally recognized in this sample, the best known of which is *Sivapithecus sivalensis* (Lydekker 1879), from localities ranging in age from 9.5 to 8.5 Ma (Kelley 2002). In addition to the characters outlined above, *Sivapithecus sivalensis* is known from a suite of cranial characters strongly indicative of pongine affinities (Pilbeam 1982). These include unfused tympanic and articular portions of the temporal bone, a posterosuperiorly directed zygomatic arch with deep temporal and zygomatic processes, vertically oriented frontal squama, supraorbital costae or rims, a narrow interorbital space, elongated nasal bones, tall, narrow orbits, wide, anteriorly oriented zygoma, narrow, pear-shaped nasal aperture, externally rotated canines, long, horizontally oriented nasoalveolar clivus that is curved along its length but flat transversely and a subnasal region with the posterior pole of the premaxilla merging into the anterior edge of the maxillary palating process to form a flat, nearly continuous subnasal floor (Figure 4.3) and a strongly concave facial profile from glabella to the base of the nasal aperture. *Sivapithecus* is also likely to have been airorhynchous (having a dorsally deflected face), as in *Pongo* (Figure 4.4). All of these and other characters are described in more detail in Ward and Pilbeam (1983), Ward and Kimbel (1983), Ward and Brown (1986), Brown and Ward (1988), Ward (1997). *Sivapithecus sivalensis* is not identical to *Pongo* in cranial morphology however, even if the similarities are striking and detailed. *Sivapithecus sivalensis* is more robust than similarly size (female) *Pongo* in features related to the masticatory apparatus, including aspects of the zygomatic and temporal bones, maxillary robusticity, and molar morphology. The molars in particular are easily distinguished from those of *Pongo* in having thicker enamel and in lacking the complex pattern of crenulations seen in unworn *Pongo* molars. However, overall the number of derived characters shared with *Pongo* is quite high (Ward 1997; Kelley 2002).

*Sivapithecus indicus* (Pilgrim 1910) is the oldest species, and if the middle Chinji fossils are included, it would range from 12.5 to 10.5 Ma (Kelley 2002). It is the smallest species, at least in terms of dental size, and appears to have a slightly shorter nasoalveolar clivus or premaxilla compared to *Sivapithecus sivalensis* (see above). *Sivapithecus parvada* (Kelley 1988) is considerably larger than the other species and is known from the Nagri formation locality Y311, about 10 Ma. *Sivapithecus parvada* males are about the dental size of female gorillas. The upper central incisors are especially long mesiodistally, the M₃ is larger than the M₂, the premolars are relatively large, and the symphysis of the mandible is very deep (Kelley 2002).
4.5.2.2 *Sivapithecus* postcrania

*Sivapithecus* postcrania have been described in many publications (Pilbeam et al. 1980, 1990; Rose 1984, 1986, 1989; Spoor et al. 1991; Richmond and Whalen 2001; Madar et al. 2002). They combine a mixture of characters, some suggesting more palmigrade postures and others suggestive of more suspensory positional behavior. This has been interpreted by some to indicate that *Sivapithecus* is more primitive than any living hominoid, all of which, even humans, share numerous characters of the shoulder and forelimbs related to suspensory behavior or an ancestry of suspensory behavior (Pilbeam 1996, 1997; Pilbeam and Young 2001, 2004). This view, also shared by McCrossin and Benefit (1997), has dramatic implications for the interpretation of the hominoid fossil record. All Late Miocene hominoids, including *Sivapithecus*, share many characters of the cranium and dentition with living hominoids. If *Sivapithecus* is more primitive than extant hominoids, given its apparently primitive postcrania, then all the apparently derived hominid features of *Sivapithecus* would have evolved in parallel in *Sivapithecus*, and as these authors suggest, by extension in all Late Miocene hominoids (*Oreopithecus*, *Dryopithecus*, *Ouranopithecus*, and *Lufengpithecus*, etc.). These parallelisms include not only craniodental morphology but also details of life history and as it turns out, many postcranial characters as well. In fact, the apparently primitive characters of *Sivapithecus* postcrania are small in number compared to the large number of clearly derived hominid characters from throughout the skeleton and known biology of all Late Miocene hominids. Rather than rejecting the hominid status of *Sivapithecus* and other Late Miocene hominids because not all of their postcranial morphology is strictly hominid or even extant hominoid-like, it is much more likely that these few characters reflect mosaic evolution of the hominid skeleton, uniquely derived features of the anatomy of *Sivapithecus*, as well as some parallelism in extant hominoids (Begun 1993; Begun et al. 1997; see below).

*Sivapithecus* postcrania, though they have been the subject of more discussions related to phylogeny, are actually less well known than *Proconsul*, *Equatorius*, or *Nacholapithecus*. The following is summarized mainly from Rose (1997), Richmond and Whalen (2001), and Madar et al. (2002). Much more information is available from all the references cited earlier. Two species of *Sivapithecus* are known from the humerus, which is unlike that of modern hominoids in the curvature of the shaft and the development of the deltopectoral plane. The bicipital groove is also broad and flat and suggests a posteriorly oriented humeral head, as in the Early and Middle Miocene Hominidea described earlier. However, the humerus has such an unusual morphology that the orientation of the head,
which is already somewhat difficult to predict from bicipital groove position (Larson 1996), cannot be reconstructed with great confidence. Nevertheless, the morphology of the proximal humerus in *Sivapithecus* is suggestive of some form of pronograde quadrupedalism as seen, for example, in extant cercopithecoids. If the humeral head were oriented more posteriorly, it would also be consistent with a scapula that is placed on the side of a compressed rib cage, as in typical mammalian quadrupeds, and unlike extant hominoids (Rose 1989; Ward 1997). No direct evidence is available for the thorax or any part of the axial skeleton of *Sivapithecus*, however, so this will also have to await further discoveries. The deltopectoral plane and the curvature of the shaft of the humeri in *Sivapithecus* are quite strongly developed compared to most cercopithecoids and indicate in my view a unique form of positional behavior that is neither extant hominoid nor extant cercopithecoid-like (Madar, et al. 2002). The distal end of the humerus is in the main hominoid-like, including a well-developed trochlea separated from the capitulum by a deep, well-defined groove (the zona conoidea), though in a few details of the posterior surface there are similarities to *Proconsul* and *Kenyapithecus* (Rose 1997). Overall, however, the functional morphology of the elbow of *Sivapithecus* is most like the hominoid elbow in its ability to resist movements other than flexion and extension at the elbow joint, a hallmark of the Hominidea (Rose 1988).

The *Sivapithecus* forearm is poorly known, especially the proximal portions of the ulna and radius, which would help to more fully understand the *Sivapithecus* elbow. A juvenile radial shaft is known that is described as *Proconsul*-like, though as a juvenile it is not clear to what extent any hominoid-like characters, such as curvature and the nature of the ligamentous/muscular insertions, would be expressed. On the other hand, the few carpal bones that are known show a mixture of hominoid and non-hominoid features. The capitate of *Sivapithecus* has a somewhat expanded and rounded head, as in great apes, but overall is transversely narrow and elongated proximodistally compared to great apes. The joint surface for the third metacarpal is irregular as in great apes. The hamate is similar in length/breadth proportions and has a less strongly projecting hamulus than do great ape hamates (Spoor et al. 1991). The joint on the hamate for the triquetrum is oriented as in gorillas and also most other nonhominoid anthropoids, and its shape suggests a stabilizing function at the wrist, which differs fundamentally from the typical mobility of the ulnar side of the wrist in extant hominoids. The proximal end of a first metacarpal is similar in morphology to hominids and Early Miocene Hominidea in being saddle shaped, a configuration considered to represent a good compromise between mobility and stability in a wide variety of positions. The manual phalanges are long and curved, with strongly developed ridges for the flexor muscles and their sheaths.
One complete phalanx has a relatively deep and somewhat dorsally positioned articular surface for the metacarpal head, which is more typical of palmigrade quadrupeds. However, it is not completely clear if this is from a hand or a foot, and if the latter is the case, a similar morphology exists in some hominoids as well.

The hindlimb of *Sivapithecus* is less well known but generally more similar to extant hominoids than the forelimb. The femur is known from proximal and distal ends but not from the same individual (Rose 1986; Madar et al. 2002). The hip joint as represented by the femoral head and neck was mobile in many directions, though it has a well-developed fovea capitis, unlike *Pongo*, which is highly distinctive (but not unique) in lacking a ligamentum teres of the femur and thus its attachment site to the femur, the fovea capitis. The distal end of the femur preserves evidence of a knee joint that is consistent with this interpretation, implying a knee joint loaded in positions away from the sagittal plane. The knee also has a number of features allowing for rotation of the leg and foot to adjust the lower extremity to a variety of positions close to and further away from the center of mass (Madar et al. 2002). In all of these features, the femur is more like that of hominoids than other anthropoids.

Other hindlimb elements include several tarsal bones, phalanges, and a well-preserved hallucial skeleton (Conroy and Rose 1983; Rose 1984, 1986, 1994). The tarsals are perhaps more like those of great apes than any other part of the post-cranium of *Sivapithecus* (Rose 1984; Madar et al. 2002), indicating the presence of a broad foot, able to assume many positions but stable in all of them, and supportive of body mass loading from many directions. The hallux or big toe is strikingly robust, much more so than in *Pongo*, and indicates a strongly developed grasping capability in the foot. The phalanges are also by and large hominoid-like, with well-developed features related to powerful flexion of the toes, a critical function in antipronograde activities (climbing as well as suspension).

### 4.5.2.3 *Sivapithecus* phylogeny and paleobiology

Overall, the morphology of *Sivapithecus* strongly supports a close phylogenetic relationship to *Pongo* but an adaptation that differed from *Pongo* in important aspects. Microwear suggests that *Sivapithecus* had a diet that was similar to that of chimpanzees while gnathic morphology suggests more of a hard object diet (Teaford and Walker 1984; Kay and Ungar 1997). Perhaps this reflects a capacity to exploit fallback or “keystone” resources in times of scarcity. Most hominoids are known to practice this strategy (Tutin and Fernandez 1993; Tutin et al. 1997). The case of gorillas may be most relevant to the question of the diet of *Sivapithecus*. 
Most gorillas have diets similar to chimpanzees but are able to exploit terrestrial herbaceous vegetation (THV) in lean seasons when soft fruit is less available or in contexts in which they are sympatric with chimpanzees (Tutin and Fernandez 1993; Tutin et al. 1997). The microwear results may reflect the preferred and most common components of the diet, while the morphology of the jaws and teeth may reflect a critical adaptation to a keystone resource on which survival would depend during stressful periods.

Postcranial evidence clearly indicates that *Sivapithecus* was not orang-‐like in its positional behavior. In fact, it was unique; there are probably no living analogues. *Sivapithecus* combines clear indications of pronograde forelimb postures and a palmigrade hand position with more antipronograde activities such as vertical climbing and clambering implied by elbow joint stability over a wide range of flexion/extension, powerful grasping hands and feet, an especially powerful hallux, and hindlimbs capable of wider ranges of joint excursions than in extant pronograde quadrupeds (Madar et al. 2002). It is difficult to imagine exactly what the positional behavior of *Sivapithecus* might have been like.

One constant in the postcranial functional morphology of *Sivapithecus* is arboreality. Perhaps *Sivapithecus* used its powerful limbs in climbing and bridging or clambering activities, spreading the limbs across multiple supports to access smaller branches. In a sense it is orang-‐like without the suspension. While *Sivapithecus* managed to distribute its considerable body mass across the tops of several branches, orangs do the same but from below. The advantage to suspension is added stability on horizontal supports in large animals, which otherwise need to generate very high levels of torque to stay atop a branch (Grand 1972, 1978; Cartmill 1985). Orang males are larger than *Sivapithecus* and may be beyond the threshold where pronograde limb postures are possible in the trees. The fact that the proximal half of the humerus of *Sivapithecus*, while similar to that of a pronograde quadruped, is exceptionally robust, with extremely well-‐developed shoulder muscle attachments, may be an indication of a unique approach to this problem.

*Sivapithecus* does share numerous postcranial features, especially of the elbow and hindlimb, with extant hominoids and *Dryopithecus*. It is therefore possible that the more monkey-‐like morphology of the proximal humerus and portions of the hands and feet are actually homoplasies with cercopithecoids caused by the adoption of more pronograde postures in a hominoid that evolved from a more suspensory ancestor (Begun et al. 1997). This requires many fewer homoplasies than the alternative hypothesis that all extant hominoid characters in all Late Miocene hominoids are homoplasies (Pilbeam 1996, 1997; McCrossin and Benefit 1997; Pilbeam and Young 2001, 2004). There is some evidence from the functional morphology of *Sivapithecus* to support the hypothesis that its
form of pronograde arboreal quadrupedalism is actually superimposed on a suspensory hominoid groundplan.

The problem of angular momentum causing instability in a large mammal standing on top of a branch is alleviated in part by spreading the limbs apart on a wide support or across several supports, which *Sivapithecus* seems to have been capable of doing (Madar et al. 2002). It can also be all alleviated by placing the center of mass closer to the support, which is suggested for *Griphopithecus* (Begun 1992a), and may have also occurred in *Sivapithecus*, as a consequence of its habitually more laterally placed limbs (Madar et al. 2002). The positioning of the limbs more laterally in hominoids is part of the suite of characters related to trunk morphology and scapular position. It is not facilitated by monkey-like trunks and scapular positions, which promote more parasagittal limb movements. Finally, an important response to angular momentum is to increase the torque generated by the limbs on the support, to prevent excessive excursions from a balanced position, especially when a single support of only modest size is used, again suggested to be an aspect of the positional behavior of *Sivapithecus* (Madar et al. 2002). Higher torque results from more powerful gripping, also characteristic of *Sivapithecus* (Madar et al. 2002), and may have been boosted by especially powerfully developed shoulder joint adduction and medial rotation, particularly if the shoulder is in a relatively abducted position to begin with, as in hominoids.

The morphology of the proximal half of the humerus in *Sivapithecus* is consistent with very powerful adduction and medial rotation by deltoideus and pectoralis major. These muscles left extremely prominent scars on the humerus of *Sivapithecus*. If the arm of *Sivapithecus* were positioned as in extant hominoids, laterally on a posteriorly positioned scapula, the adduction and medial rotation capacities of deltoideus and pectoralis would be increased by increasing the relative mass of the clavicular portion of deltoideus, which is unknown in *Sivapithecus*. However, in addition to the need for powerful muscles, which existed in *Sivapithecus*, these functions would also be enhanced by other attributes known in *Sivapithecus*. The extension of the muscles along the shaft distally and the possible decreased humeral neck torsion would increase the moment arm for these muscles in adduction and reposition the insertions of these muscles medially, possibly to make more of the deltoideus available for adduction and medial rotation. The strong mediolateral curvature of the shaft may result from high mediolateral bending stresses that would result from very powerful shoulder adduction on a fixed limb. The proximal shaft is also very broad mediolaterally, suggestive of strong mediolaterally directed stresses. While speculative, it is certainly conceivable that the upper part of the forelimb of *Sivapithecus* was less monkey-like than generally perceived and may not imply that much of the trunk
was of the primitive anthropoid type. A small number of autapomorphies of the shoulder of *Sivapithecus* may have allowed this taxon to practice a relatively efficient form of arboreal pronograde quadrupedalism while maintaining the capacity for many of the antipronograde activities of hominoids, though probably not frequent upper limb below branch suspension. This hypothesis is functionally consistent with the morphology of the *Sivapithecus* postcranial in general and is certainly more parsimonious than the hypothesis that would interpret all Late Miocene hominoid characters as homoplasies.

### 4.5.3 *Ankarapithecus*

Aside from some material from Nepal attributed to *Sivapithecus* (Munthe et al. 1983), *Sivapithecus* is known only from India and Pakistan. Specimens from central Anatolia, Turkey, once attributed to *Sivapithecus* (Andrews and Tekkaya 1980) are now assigned to *Ankarapithecus*, following the original conclusions of Ozansoy (1957, 1965). *Ankarapithecus meteai* is known from a male palate and mandible from two different individuals and a female partial skull (mandible and face) from a third locality, all close to each other in location and geologic time (Kappelman et al. 2003a). Begun and Güleç (1995, 1998) resurrected the nomen *Ankarapithecus* based mainly on the morphology of the premaxilla and the relationship between this bone and the maxilla but concluded that *Ankarapithecus* is nonetheless in the pongine clade. Alpagut et al. (1996) and Kappelman et al. (2003) described newer and much more complete fossils of *Ankarapithecus* and concluded that it is a stem hominid (sharing a common ancestor with both pongines and hominines). The new fossils discovered and described by these authors include the region around the orbits, which lacks some of the characters of *Sivapithecus*. The interorbital region is intermediate in breadth between *Pongo*, with the narrowest interorbitals, and African apes and the orbits themselves are broad rather than tall and narrow. Alpagut et al. (1996) and Kappelman et al. (2003) also interpret the supraorbital region as a supraorbital torus, characteristic of African apes and humans and some European Late Miocene taxa, and they interpret a frontal sinus in the cranium of *Ankarapithecus* as a frontoethmoidal sinus. These authors see the mixture of hominine and pongine characters as an indication that *Ankarapithecus* precedes their divergence.

#### 4.5.3.1 Craniodental evidence

The frontal sinuses in *Ankarapithecus* appear to be confined to the frontal squama and do not invade the frontal supraorbital region from a broad expansion of the
Ethmoidal sinuses. They are positioned and developed as in extant taxa with frontal pneumatizations derived either from the sphenoidal or maxillary sinuses and are unlike those derived from the ethmoid (see above discussion). The frontal pneumatization in *Ankarapithecus* is unlikely to be a frontoethmoid sinus and thus is not a synapomorphy of the hominines. The supraorbital region, while robust, is morphologically similar to the robust supraorbital costae of large orangs or *Cebus* and also unlike the bar-like supraorbital tori of African apes. Thus, the supraorbital region of the *Ankarapithecus* cranium is more pongine than hominine-like and, like the maxilla, probably represents the primitive morphology for the pongines (see below).

In *Ankarapithecus* the premaxilla, the portion of the palate with the alveoli for the incisors and the mesial half of the canines, is unlike that of *Sivapithecus* and more like that of African apes. The premaxilla is long, but it does not overlap the palatine process of the maxilla to fill the incisive fossa to the degree seen in *Sivapithecus*. Instead, the subnasal fossa is stepped (there is a drop between the base of the nasal aperture and the floor of the nasal cavity) into an incisive fossa that is most like that of some chimpanzees, a relatively large depression opening into a canal (the incisive canal that runs between the premaxilla and maxilla to exit on the palatal side via the incisive foramen). *Dryopithecus* has a similar configuration of the subnasal fossa, incisive canal and incisive foramen, though the fossa is larger and the canal is shorter in length and larger in caliber, as in some gorillas (Begun 1994a) (Figure 4.3). The premaxilla of *Ankarapithecus* is curved or convex anteroposteriorly as in *Sivapithecus*, African apes, and *Dryopithecus* but it is also convex transversely, as in African apes and *Dryopithecus* unlike *Sivapithecus*, which has a transversely flatter premaxilla. In all of these features, *Ankarapithecus* expresses a condition intermediate between pongines and hominines, which I consider primitive for the pongines (Begun and Güleç 1998). Alpagut et al. (1996) and Kappelman et al. (2003b) have suggested that these characters indicate that *Ankarapithecus* precedes the divergence of pongines and hominines. Other features of the morphology of *Ankarapithecus* resemble pongines more clearly, including canine implantation, zygoma size and orientation, orbital margin morphology, nasal length, and dental morphology. Overall, *Ankarapithecus* most closely resembles *Sivapithecus* and *Pongo* but retains a more primitive palatal morphology that suggests it is at the base of the pongine clade.

### 4.5.3.2 Postcranial evidence

Some postcrania of *Ankarapithecus* are known, including a well-preserved radius and two phalanges (Kappelman et al. 2003). A femur tentatively identified as
primate is more likely in my opinion to be from a carnivore. The radius shares characters with extant great apes including features of the radial head and comparatively long radial neck (Kappelman et al. 2003). Other hominoid-like features described or figured in Kappelman et al. (2003) but not identified as hominoid-like by these authors include a proximodistally compressed and more circular radial head, a deep radial fovea, flat as opposed to concave shaft surface along the anterior surface, a more distal origin of the interrosseus crest, and a smooth distal dorsal surface. The specimen actually strikes me as quite hominoid-like with a few features more normally associated with large nonhominoid anthropoids, a pattern more or less in keeping with other Late Miocene hominoids. The phalanges are said to be relatively straight and thus non-hominoid-like, but only distal portions are preserved. They too strike me as more hominoid-like than Kappelman et al. (2003) suggest, given its distal shaft robusticity, dorso-palmar compression, and distopalmarly projected condyles. The curvature and ridges for the flexor musculature are said to be poorly developed compared to hominoids, but this may be related to many factors (preservation, age, digit attribution).

Overall, *Ankarapithecus* is characterized by many features found in other pongines and is probably the most basal known member of that clade. Like *Sivapithecus*, it was much more massive in the development of its masticatory apparatus than *Pongo*, and its postcranium, though very poorly known, suggests arboreality and at least some features of hominoid-like antipronograde positional behaviors, but probably lacking the degree of suspension seen in *Dryopithecus*, *Oreopithecus*, and extant hominoids (but see Kappelman et al. 2003).

4.5.4 Other probable fossil pongines

4.5.4.1 *Gigantopithecus*

Extremely large fossil hominoids, larger than any extant primate, have been known from Asia since the early part of the twentieth century. *Gigantopithecus blacki* (Koenigswald 1935) is a Pleistocene taxon known from numerous isolated teeth and a few mandibles. It is recent enough to be outside the purview of this review and has been described many times elsewhere. *Gigantopithecus giganteus* (Pilgrim 1915) is a Late Miocene possible member of this genus known from a lower molar and mandible from the Late Miocene of the Siwaliks that are mainly distinguished from *Sivapithecus* by size. They are larger than any *Sivapithecus* and the only known mandible is distinctive in having reduced anterior tooth crown heights and molarized or enlarged premolars. *G. blacki*, on the other hand, is
larger still and has highly complicated postcanine occlusal morphology and relatively even larger mandibles and smaller anterior teeth. Because *G. giganteus* and *G. blacki* share characters of the lower jaws and teeth that appear commonly during the course of hominoid evolution (in fact, in many other mammal lineages as well), the relationship between the two is uncertain. Jaws and teeth in general, and mandibles in particular, are magnets for homoplasy in primate evolution (Begun 1994b), and this may be another example. However, the most parsimonious hypothesis is that *G. giganteus* is a primitive member of the *Gigantopithecus* clade and that the strong similarity to *Sivapithecus* in the molars, apart from size, suggests that it is the sister clade to that taxon. If a better ancestor for *Gigantopithecus blacki* turns up, the nomen *Indopithecus* is available for *G. giganteus* (Table 4.1).

### 4.5.4.2 Lufengpithecus

Thousands of fossils, mostly isolated teeth, are known from a number of localities in Yunnan province, southern China. These are attributed to the genus *Lufengpithecus* (Wu 1987), widely believed to be a pongine. Kelley tentatively recognizes three species of *Lufengpithecus*, distinguished mainly by size and geography (each is known from a single site). *Lufengpithecus* shares a few cranial characters with other pongines including a small, pear-shaped nasal fossa, aspects of the implantation of the canine roots in the maxilla, a deep canine fossa, supraorbital costae, and anteriorly oriented zygoma (Schwartz 1997 and personal observations). However, while it lacks many of the detailed similarities of the face between *Sivapithecus* and *Pongo*, its teeth are much more like those of *Pongo* than *Sivapithecus* in details of occlusal morphology, including the unusual presence of highly complex wrinkling or crenulations (Kelley 2002). The face of *Lufengpithecus* is unlike those of *Sivapithecus* and *Pongo* in having broad orbits, a broad interorbital space, a comparatively short premaxilla, high crowned incisors and canines, compressed and very tall crowned male lower canines. Though very damaged, my impression is that the nasal floor is unlike the smooth floor of *Sivapithecus* and *Pongo* but possibly more similar to the morphology in *Ankarapithecus*.

* L. *lufengensis* (Xu et al. 1978) from a site near Shihuiiba, in Lufeng county, is the best-known species of the genus and is also known from a number of postcranial remains including fragments of a scapula, clavicle, radius, first metatarsal, and two phalanges. None have been published in detail, but all of these specimens show clear indications of modern hominid morphology associated with suspensory positional behaviors (personal observations). This is especially true of the phalanges, which are strongly curved and bear the markings of
powerful flexor tendons. The metatarsal is similar to that of *Sivapithecus* in its relative robusticity.

Until the fossils attributed to *Lufengpithecus* are published in detail, it will be impossible to be confident in assessing their taxonomic and phylogenetic relations. At this point, it seems likely that *Lufengpithecus* is a pongine and probably a sister taxon to the *Pongo–Sivapithecus–Ankarapithecus* clade. However, the generally more *Pongo*-like morphology of the molar occlusal surfaces and the more clearly hominoid-like morphology of the postcrania are enigmatic and suggest caution in interpreting evolutionary relationships.

### 4.5.4.3 *Khoratpithecus*

Two samples of East Asian Miocene hominoids have recently come to light. Chaimanee et al. (2003) describe a Middle Miocene sample of hominoids from Thailand they originally attributed to a species of *Lufengpithecus*. The age of the locality is not completely certain, and it is possible that the fauna could be correlated with a more recent magnetostratigraphic interval, but for now the sample is considered to date to the late Middle Miocene or early Late Miocene. However, on the basis of more recently discovered Late Miocene hominoid fossils from Thailand, Chaimanee et al. (2004) described a new genus, *Khoratpithecus*, and revised their previous taxonomic conclusions to include the Middle Miocene taxon in the newly named genus as well. The fossils, a sample of isolated teeth and a well-preserved mandible, are very similar to *Lufengpithecus* but show a number of differences in the anterior dentition and lower jaw (Chaimanee et al. 2004). Chaimanee et al. (2004) interpret *Khoratpithecus* to be more closely related to *Pongo* than is any other pongine, mainly based on the shared derived character of a missing anterior digastric muscle. More fossils are needed to test this hypothesis more fully. The greater significance of these discoveries is the location, in Thailand, and the possibly early age, Middle Miocene.

### 4.5.5 Fossil hominines

At about the same time that hominids appear in Asia, they make their first appearance in Europe. As is the case with the earliest pongines, the earliest hominines lack a number of synapomorphies of living hominines (African apes and humans) and are less distinct from related non-hominines than are more recently evolved hominines. This has led naturally to differences of opinion
regarding the systematics of this group. There are three main interpretations of the evolutionary relations among the taxa included here in the Homininae. As noted, some researchers conclude that no known Eurasian Late Miocene taxon has a specific relationship to extant hominoids (Pilbeam 1996, 1997; McCrossin and Benefit 1997; Pilbeam and Young 2001, 2004). Most researchers, however, accept the hominid status of these fossil taxa but are divided as to their inter-relationships. Some researchers (Andrews 1992) have concluded that European Late Miocene hominids are best viewed as stem hominids, preceding the divergence of hominines and pongines. Others interpret most or all Eurasian hominines to be members of the pongine clade (Moya-Sola et al. 1995). Finally, some researchers interpret most or all European hominids to be hominines, although there is disagreement among them as to the precise pattern of relations (Bonis and Koufos 1997; Begun and Kordos 1997). As my interpretation falls with the last group, this will be reflected in this chapter. I will however attempt to outline the major arguments from each perspective.

4.5.6 Pierolapithecus

The oldest known genus of hominine is the recently described partial skeleton of *Pierolapithecus* (Moya-Solà et al. 2004). The specimen, from northern Spain, is dated to about 13 Ma based on biostratigraphic evidence (Moya-Solà et al. 2004). The specimen includes most of a face which, though badly distorted, preserves nearly all the teeth and many informative facial characters. It also includes a partial postcranial skeleton, the most informative parts of which are some lumbar vertebrae, ribs, and a number of hand and foot bones. Moyà-Solà et al. (2004) interpret *Pierolapithecus* to be a basal or stem hominid. They cite the lumbar vertebrae, which preserve evidence of a hominoid-like vertebral column and by extension rib cage. This is indicated by the position of the lumbar transverse processes, placed more posteriorly in hominoids to stiffen the lower back (Ward 1993). However, the morphology of the lumbar vertebrae in *Pierolapithecus* is more hylobatid-like, extant hominids having even more posteriorly positioned transverse processes. Other aspects of the postcranium that clearly support the hominoid status of *Pierolapithecus* include ribs indicative of a broad, anteroposteriorly compressed rib cage, robust clavicle, and a wrist morphology indicating no direct contact between the carpus at the triquetrum and the ulna (Moyà-Solà et al. 2004). These features are shared with all extant hominoids, though hylobatids are somewhat intermediate in carpal/ulnar contact between non-hominoid anthropoids with direct contact and hominoids with a large intervening articular meniscus (Lewis 1989).
The carpals in general are hominid-like in their overall morphology, including relative size, robusticity, and general pattern of the orientation of the joint surfaces. The lunate, triquetrum, and hamate in particular closely resemble small chimpanzees, but it is not clear if these are derived characters for hominines or hominids. Moyà-Sola et al. (2004) describe the phalanges of Pierolapithecus as being relatively shorter, less curved, and with metacarpal joint surfaces facing more dorsally than in Dryopithecus, a clearly suspensory hominine, which they interpret to mean that Pierolapithecus had a palmigrade hand posture. At the same time, the attributes of the thorax and hand suggest antipronograde limb positions, which is somewhat contradictory. They resolve this dilemma with the suggestion that Pierolapithecus was a powerful vertical climber but not suspensory. This is similar to the suggestion made earlier regarding Sivapithecus, though the morphology of the phalanges does not in fact rule out suspension. The phalanges are curved compared to most arboreal primates and have strongly developed flexor muscle attachments, even if these are not so strongly expressed as in Dryopithecus (Begun and Ward 2005).

Moyà-Sola et al. (2004) interpret various craniodental attributes of Pierolapithecus to reflect its stem hominid status as well. The face is prognathic with an enlarged premaxilla. The zygomatic root is high, the nasal aperture broad, and the postcanine teeth have a typical hominid morphology (elongated, relatively large M1, absence of cingula, reduced premolar cusp heteromorphy, buccolingually large incisors, compressed canines). The premaxilla is expanded compared to early and exclusively Middle Miocene hominoids and appears to have an overlap posteriorly with the maxilla as in hominids. However, according to these authors, it lacks the distinctive attributes of either the hominine of pongine clade.

Some of the distinctive attributes of Pierolapithecus may be related to distortion. The glabella is unlikely to have been as posterior as it appears, the midface is clearly badly damaged and was not as prognathic as in Afropithecus, as the authors suggest, and the premaxilla is obviously displaced relative to the palatine process of the maxilla (Begun and Ward 2005). In my view, the face much more closely resembled Dryopithecus, though it is still distinct enough to justify a separate genus.

All in all, Pierolapithecus closely resembles Dryopithecus, known from contemporaneous and younger localities in Spain and elsewhere in Europe. In fact, it may be that Pierolapithecus catalaunicus is synonymous with Middle Miocene Dryopithecus fontani (see below). Both are from the same time period and nearby localities, but the mandibles and humeral shaft of D. fontaini cannot be directly compared to the sample of Pierolapithecus. If future discoveries of fossils of both samples reveal that they are synonymous, Dryopithecus would have priority over
Pierolapithecus. Younger samples currently attributed to Dryopithecus, which are distinguishable from Pierolapithecus, would require another nomen, several of which are available. Once again, it is premature to revise the taxonomy of Dryopithecus in the absence of more data from the Middle Miocene samples.

While aspects of the postcrania may be more primitive, and aspects of the face unique, justifying a separate genus from Late Miocene Dryopithecus, there is evidence to suggest that Pierolapithecus is a stem hominine and not as Moyà-Solà et al. (2004) conclude a stem hominin (Begun and Ward 2005). Details of dental morphology are strikingly similar to Dryopithecus. Despite the unusually small M3 and elongated upper canine, most of the teeth could easily be mistaken for those of Dryopithecus and show features distinctive for that genus including relatively tall crowned and mesiodistally narrow upper incisors, compressed canines, premolars with prominent cusps separated by a broad deep basin and molars with marginalized or peripheralized, relatively sharp cusps. The contact between the premaxilla and the maxilla appears to also have been very similar to Dryopithecus in being stepped with only a modest degree of overlap between the two. The supraorbital region, though described by Moyà-Solà et al. (2004) as having thin supraorbital arches, actually closely resembles Dryopithecus specimens from Spain and Hungary, with subtle tori emerging from a more prominent glabella. In my view, Pierolapithecus is close to the common ancestor of the Hominidae but already shares a common ancestor with the Homininae (Begun and Ward 2005). Its postcranial morphology, however, is probably very close to that of the hominid ancestral morphotype (Figure 4.2).

4.5.7 Dryopithecus

The first fossil hominoid genus ever described was Dryopithecus, based on fossils from St. Gaudens in France, attributed to D. fontani (Lartet 1856). Today Dryopithecus is known from at least four species ranging in time from the late Middle Miocene, about 12–13 Ma, to the Late Miocene, about 9 Ma, and ranging geographically from Spain in the west to Georgia in the east (a mandible from Gansu province in China with badly damaged teeth attributed to Dryopithecus wuduensis [Xu and Delson 1988] is not distinguishable in my view from a large cercopithecoid). The following account is mainly from Begun (2002). The oldest species of Dryopithecus, D. fontani, is known from Middle Miocene deposits in France and Austria and is roughly contemporaneous with Pierolapithecus, and Sivapithecus. D. crusafonti is known from two localities in Spain from the early part of the Late Miocene. D. laietanus and D. brancoi are the latest occurring species of Dryopithecus and are known respectively from sites in Spain for the former and
Germany and Hungary for the latter. Two younger Miocene teeth from Udabno, in Georgia, are often attributed to *D. brancoi* but may be another species.

### 4.5.7.1 *Dryopithecus fontani*

*D. fontani* is known from three male mandibles and a humerus, all from the same locality in France, and a female mandible from Austria. Two isolated upper teeth from La Grive in France usually attributed to *D. fontani* may belong to *Pierolapithecus*. They are the only specimens currently attributed to *D. fontani* that can be directly compared to *Pierolapithecus* (Begun et al. 2006). The mandibles and their dentitions are typically hominid in being comparatively robust with well-developed symphyseal tori, large incisors, compressed canines, elongated post-canine teeth with peripheralized cusps and lacking cingula, P4 with trigonids and talonids of nearly equal height, and molars of nearly equal size, especially M1 and M2. The teeth of all *Dryopithecus* are thinly enameled with dentine horns penetrating well into the enamel caps. *D. fontani* is distinguished from other species of *Dryopithecus* in having a mandible that shallows (becomes lower compared to breath) distally, a high frequency of buccal notches on the lower molars, and comparatively robust lower canines.

*D. fontani* is also known from a humeral shaft from the type locality that has been described as chimpanzee-like (Pilbeam and Simons 1971; Begun 1992a). It is the only nearly complete humerus of the genus. It is comparatively long and slender with poorly developed muscle insertion scars and a slight mediolateral and anteroposterior curvature. Neither the proximal nor the distal epiphyses are preserved, but the diaphysis preserved close to each epiphysis is hominoid-like. Proximally, it is rounded in cross section with a bicipital groove position suggesting some degree of humeral torsion (but see Rose 1997; Larson 1998). Distally, it is mediolaterally broad and anteroposteriorly quite flat, with a large, broad, relatively shallow olecranon fossa (Begun 1992a).

### 4.5.7.2 Spanish *Dryopithecus*

*D. crusafonti* (Begun 1992c) is known from a sample of isolated teeth and a palatal fragment from one site and a well-preserved mandible from a second, both in northern Spain. *D. crusafonti* is dentally similar to *D. fontani* but has distinctive upper central incisors, a more robust mandible lacking the distal shallowing, upper molars of nearly the same size and a number of subtle features of dental morphology.
D. laietanus (Villalta and Crusafont 1944) is known from several slightly younger sites in Spain. Dentally, it is smaller but similar to other species of Dryopithecus and lacks the unique dental characters of D. crusafonti. It is the best-known species of the genus because of the recent discovery of a partial skeleton (Moya-Sola et al. 1996). Like D. crusafonti, D. laietanus has tall, relatively narrow upper central incisors, though not to the degree seen in D. crusafonti. The mandible is relatively robust. A partial cranium of D. laietanus displays numerous hominid cranial characters (broad nasal aperture base, high zygomatic roots, shallow subarcuate fossa, and probable enlarged premaxilla with maxillary overlap, although the specimen is damaged in that area). A few hominine characters are found on this specimen as well (supraorbital tori connected to glabella, frontoethmoidal sinus, inclined frontal squama and thin enamel).

The most significant specimen of D. laietanus is a partial skeleton that may or may not be associated with the face (Moya-Sola et al. believe they are, but the cranium was found widely separated from the postcranial specimens and appears a bit too large for the face). Nevertheless, this is an exceptional and important specimen. The most significant features of the postcranial skeleton of D. laietanus are the numerous and unambiguous indications of both well-developed suspensory positional behavior and clear hominid synapomorphies. These include elongated forelimb, large hands with powerful, curved, elongated digits, comparatively short and robust hindlimb, and a hominid-like lumbar region. Other attributes interpreted to be present in this partial skeleton, such as an elongated clavicle and limb proportions approaching those of Pongo, are based on fragmentary evidence and are less reliable. The specimen has some unusual features for a hominid such as short metacarpals, but overall it is quite modern. The humerus, though fragmentary, is like that of D. fontani and unlike that of Sivapithecus.

4.5.7.3 Dryopithecus brancoi

In the second half of the nineteenth century, shortly after the initial discovery and description of Dryopithecus fontani, additional fossil hominoid teeth began to turn up in Germany. These were eventually assembled to define the new species, D. brancoi (Schlosser 1901), though not before considerable taxonomic shuffling (see Begun 2002 for more historical details). D. brancoi is based on an isolated M3, which, while not the ideal type specimen, can be effectively distinguished from the other species. To help in species identification, the species diagnosis was revised by Begun and Kordos (1993) based on the excellent sample from Rudabanya, Hungary, and this is the definition used here.
D. brancoi is known the type locality in Germany (Salmendingen) and from Rudabánya and is dated between 9 and 10 Ma. The younger date is based on the type locality, which is poorly dated, such that the older age may be more accurate. Six other localities in Germany, Austria, and Georgia may also contain D. brancoi, but as the specimens are all isolated teeth it is difficult to be certain. D. brancoi shares all the hominid characters already described for other Late Miocene hominids, but the cranium is better preserved in this taxon than in any other, and provides additional details (Table 4.3).

Table 4.3
Great ape and African ape craniodental character states of Dryopithecus

<table>
<thead>
<tr>
<th>Great ape character states</th>
<th>African ape character states</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labioliungually thick incisors</td>
<td>Biconvex premaxilla</td>
</tr>
<tr>
<td>Compressed canines</td>
<td>Stepped subnasal fossa</td>
</tr>
<tr>
<td>Elongated premolars and molars</td>
<td>Patent incisive canals</td>
</tr>
<tr>
<td>M1 = M2</td>
<td>Broad, flat nasal aperture base</td>
</tr>
<tr>
<td>No molar cingula</td>
<td>Shallow canine fossa</td>
</tr>
<tr>
<td>Reduced premolar cusp heteromorphy</td>
<td>Supraorbital torus</td>
</tr>
<tr>
<td>High root of the zygomatic</td>
<td>Inflated glabella</td>
</tr>
<tr>
<td>Elongated midface</td>
<td>Frontal sinus above and below nasion</td>
</tr>
<tr>
<td>Broad nasal aperture below the orbits</td>
<td>Projecting entoglenoid process</td>
</tr>
<tr>
<td>Reduced midfacial prognathism</td>
<td>Fused articular and tympanic temporal</td>
</tr>
<tr>
<td>Elongated, robust premaxilla</td>
<td>Broad temporal fossa</td>
</tr>
<tr>
<td>Premaxilla–palatine overlap</td>
<td>Deep glenoid fossa</td>
</tr>
<tr>
<td>Shallow subarcuate fossa</td>
<td>Elongated neurocranium</td>
</tr>
<tr>
<td>Enlarged semicircular canals</td>
<td>Moderate alveolar prognathism</td>
</tr>
<tr>
<td>Large brain</td>
<td>Klinorhynchy</td>
</tr>
<tr>
<td>High cranial base (Begun 2004)</td>
<td></td>
</tr>
</tbody>
</table>

D. brancoi shares with other Dryopithecus all the details of canine and postcanine tooth morphology outlined above. It shares relatively narrow and labioliungually thick upper central incisors with other Dryopithecus, though not to the degree seen in D. crusafonti. In addition to the hominid characters of Dryopithecus previously noted (Table 4.3), D. brancoi preserves a few details of the face and many details of the neuro and basicrania, with further evidence of its hominid status. The zygoma are high, prominent, and oriented anterolaterally, as in hominines, and the number and position of the zygomaticofacial foramina is variable (this character has been proposed as one that could establish the pongine affinities of Dryopithecus, but the configuration in D. brancoi is hominine-like [Kordos and Begun 2001]). The neurocranium is large with a reconstructed cranial capacity in the range of extant chimpanzees (D. brancoi is the only Late Miocene hominid for which cranial capacity reconstruction is possible from
direct measurements of the brain case [in two individuals]) (Kordos and Begun 1998, 2001; Begun and Kordos 2004).

Among the hominine characters preserved in the cranial sample of *D. brancoi* are a relatively low and elongated neurocranium, with the inion displaced inferiorly (Table 4.3). The interorbital and supraorbital regions have sinuses that are largest above and Glabella is prominent and continuous with small supraorbital tori separated from the frontal squama by a mild supratoral sulcus (Begun 1994a). The temporal bone, in addition to preserving evidence of a shallow subarcuate fossa (a hominid character), suggests fusion of the articular and tympanic portions and preserves details of the temporomandibular joint found only in hominines (Kordos and Begun 1997). A recently discovered and described cranium is the first to include a well-preserved neurocranial and facial skeleton in connection and shows clearly that the cranium of *Dryopithecus* was klinorhynchy (having a ventrally deflected face), which it shares with African apes among the hominoids (Kordos and Begun 2001) (Figure 4.4).

The nasoalveolar clivus or premaxilla is hominine like in its orientation, size, surface anatomy, and relations (Figure 4.3). It is biconvex, long compared to Early Miocene Hominidea and hylobatids (proportionally equal in length to *Gorilla*), with a posterior pole that is elevated relative to the nasal floor, giving a stepped morphology to the subnasal fossa (Begun 1994a; Figure 4.3). The resulting incisive fossa of the subnasal floor is deep and well defined, the incisive canal is short and large in caliber, and the incisive foramen on the palatal side is comparatively large. This suite of characters is found in *Gorilla* as well and suggests that this is the ancestral morphology for hominines. *Pan* and *Australopithecus* share the synapomorphic condition of a more elongated but still biconvex premaxilla, which along with their spatulate upper central incisors and neurocranial morphology is among the most important morphological synapomorphies of the chimpanzee–human clade (Begun 1992c).

*D. brancoi* is well represented by postcrania, including a distal humerus that is hominid-like in all details related to trochlear and capitular morphology, even more so than *Sivapithecus* in having broad and shallow fossae for the processes of the radius and ulnae (see above). The ulna is robust with a strongly developed trochlear keel and a radial facet orientation that indicates forearm bones positioned for enhanced antipronograde postures (Begun 1992a). The scaphoid is *Pongo-like* in morphology and was not fused to the os centrale, as it is in African apes and humans (Begun et al. 2003). The capitate is large with a complex metacarpal articular surface, as in African apes, but the head is comparatively narrow and the bone overall is elongated compared to African apes, again more like the condition in *Pongo* and *Sivapithecus*. The phalanges are long, strongly curved, and marked by sharp ridges for the flexor musculature, indicative of
suspenory positional behavior (Begun 1993). Recently discovered femora of a very small individual of *D. brancoi* are short, with a large head, long neck, and extremely robust shaft, consistent with the hominoid pattern, and again especially similar to *Pongo*. The foot is also apelike in its broad, flat talar body and mobile but large entocuneiform and hallux.

### 4.5.7.4 Paleobiology of *Dryopithecus*

All four species of *Dryopithecus* display dental morphological characters that are very similar to extant *Pan* and suggest a soft fruit diet (Begun 1994a). Microwear analyses support this assessment (Kay and Ungar 1997). The gnathic morphology of *Dryopithecus* is gracile compared to many other Late Miocene hominids (less robust mandibles, thinner occlusal enamel, smaller attachment sites for the muscles of mastication), which is both consistent with a soft fruit diet and more similar to extant African apes, *Pan* in particular. Postcranially, *Dryopithecus* is unambiguously suspensory, but it does lack a few synapomorphies, particularly of the extremities that characterize all extant hominids. These have to do mainly with the robusticity of the bones of the carpus and tarsus, which may be attributable to a “red queen hypothesis” phenomenon, as in the case of the progressive development of shearing quotients during the course of hominoid evolution (Kay and Ungar 1997). In the fossil record of many mammals, there is evidence of a shift toward a certain adaptation (folivory, frugivory, suspension, climbing, bipedalism, etc.) that becomes increasingly refined in individual lineages descended from the common ancestor initially expressing the behavior. In order to remain competitive, the descendants must, in essence, run to stay in the same place, as increasingly efficient versions of the same adaptation appear independently (van Valen 1973). *Dryopithecus* was an arboreal, suspensory, soft fruit frugivore with a dentition similar to *Pan*, living in subtropical forests but probably capable of exploiting a variety of resources, possibly including meat (Kordos and Begun 2002).

### 4.5.8 *Ouranopithecus*

A large hominid sharing characters of *Dryopithecus* and *Sivapithecus* was first described from northern Greece and attributed to the genus *Dryopithecus* (to which *Sivapithecus* was also attributed at the time) (Bonis et al. 1975). Soon it became clear that the sample from Greece was distinct from both *Sivapithecus* and *Dryopithecus*, and the new nomen *Ouranopithecus* was proposed (Bonis and Melentis 1977).
Ouranopithecus is a large hominid, the approximate size of a large male chimpanzee or female gorilla, whose morphology is similar to that of Dryopithecus but with a much more robust masticatory adaptation (Begun and Kordos 1997). Ouranopithecus has a palate that is similar to Dryopithecus in the degree and pattern of overlap of the maxilla and premaxilla (Bonis and Melentis 1987; Begun and Kordos 1997). The morphology of the nasoalveolar clivus is also similar to Dryopithecus and extant hominines. The nasal aperture is broad at its base, the interorbital space is broad and the orbits are rectangular. The zygomatic roots arise relatively low and anteriorly on the maxilla, which is interpreted as a homoplasy with Early Miocene taxa, as a similar condition is also found in robust australopithecines that share with Ouranopithecus a very robust masticatory apparatus (Begun and Kordos 1997). The glabella is projecting, and like Dryopithecus it is continuous with subtle tori above each orbit. The frontal squama is concave above glabella, but this is somewhat exaggerated by damage. Dentally, Ouranopithecus is similar to Dryopithecus and other hominids in tooth proportions and overall dental morphology. It differs from Dryopithecus in having hyperthick occlusal enamel, molars with broad cusps and flat basins, mesiodistally longer incisors and relatively low crowned male upper canines. The mandibles are also more robust than in Dryopithecus and have strongly reinforced symphyses. The female mandibles tend to be more robust (or shallower) than the male mandibles. One mandible preserves the condylar process, which is large and strongly convex anteroposteriorly. Ouranopithecus is also known from two unpublished phalanges.

In many publications, summarized in Bonis and Koufos (1997), it has been argued that Ouranopithecus is a hominin (specifically related to humans), mainly on the basis of canine reduction and masticatory robusticity. However, these features occur repeatedly during hominoid evolution. Ouranopithecus is most parsimoniously interpreted as a terminal member of the Dryopithecus clade, with a number of craniodental specializations related to an increase in masticatory robusticity (Begun and Kordos 1997; Begun 2001, 2002). The large jaws and teeth and hyperthick enamel, as well as microwear studies, suggest an ability to exploit hard and/or tough fruits, nuts, and other dietary resources (Kay 1981; Ungar 1996; Bonis and Koufos 1997; Kay and Ungar 1997).

4.5.8.1 Graecopithecus

Another taxon, Graecopithecus (Koenigswald 1972), is also known from Greece but from a much younger locality over 200 km from the Ouranopithecus localities. It is similar to Ouranopithecus, and some have suggested that the two
samples belong to the same genus, which would be called *Graecopithecus*, since this nomen has priority (Martin and Andrews 1984). In my view the generic distinction is warranted. *Graecopithecus*, known only from a poorly preserved mandible with a fragmentary M$_1$, a very worn M$_2$ and root fragments, is similar to *Ouranopithecus* in apparently having thick occlusal enamel. However, it is the overall size of female *Ouranopithecus* but has an M$_2$ bigger than some male *Ouranopithecus*, and the M$_2$ is actually broader than the mandibular corpus at the level of the M$_2$. The symphysis is more vertical and the M$_1$ is relatively small (Begun 2002). *Graecopithecus* is morphologically distinguishable from *Ouranopithecus*, much younger in age, and geographically distant from *Ouranopithecus* localities.

4.5.9 *Oreopithecus*

The other European Miocene hominoid discovered and described during the nineteenth century is the highly unusual *Oreopithecus* (Gervais 1872). Over the years *Oreopithecus* has been called a pig, prosimian, monkey, and ape, the last being the attribution most researchers agree on today (Harrison and Rook 1997; Begun 2002). *Oreopithecus* is younger than other Late Miocene European hominoids and is known from about 6 to 7 Ma localities in Italy. At the time, most of the Italian peninsula was separated from the rest of Europe by the sea, as is today the Italian island of Sardinia, where one *Oreopithecus* locality is found. In the Late Miocene, all *Oreopithecus* localities were insular, and the faunas associated with them are unique and difficult to compare to continental European faunas (Harrison and Rook 1997). *Oreopithecus* is a product of its insular environment as well and is characterized by many unique adaptations that make it difficult to understand its relations to other hominoids.

In its craniodental morphology, *Oreopithecus* is similar to *Dryopithecus* and African apes in having apparently thin enamel, but otherwise the morphology of the teeth is quite unique. Like other hominids, *Oreopithecus* has compressed canines, reduced premolar cusp heteromorphy, and reduced or absent molar cingula. However, the incisors are small and low crowned, the P$_4$ has a primitive looking low talonid compared to the trigonid, the postcanine dentition has tall, isolated cusps, and the lower molars have a unique occlusal morphology with a centroconid connected to the four principal cusps by a well-developed system of crests. The upper molars are also strongly “cristodont,” which makes them appear similar to the lower molars, superficially resembling the condition of upper and lower molar bilophodonty in Old World monkeys.
The mandible is strongly built with some specimens being quite robust transversely and others deeper. The ramus is expansive to accommodate large temporalis and masseter muscles, which is also evidenced by the prominent temporal crests and pronounced postorbital constriction. The face is badly damaged but appears to have had a short and relatively gracile premaxilla, which is consistent with the small incisors. The brain case is also badly damaged but was clearly small, housing a much smaller brain than great apes of comparable body mass (Harrison 1989; Begun and Kordos 2004). Like *Sivapithecus* and non-hominids, the articular and temporal portions of the temporal bone are not fused but like hominids the subarcuate fossae are small. The ectocranial crests are very strongly marked while the frontal is comparatively smooth, without tori, and the postorbital constriction is marked.

The most impressive aspect of *Oreopithecus* is its postcranium. A remarkably complete but crushed skeleton along with many other isolated postcranial elements is known from *Oreopithecus*. The axial skeleton (rib cage and trunk) is hominoid-like in its short lower back and broad thorax, and the pelvis is also comparatively short and broad, as in hominids. The forelimbs are very elongated compared to the hindlimbs, the glenoid fossa of the scapula is deep, and the elbow has all the typical hominoid features described previously including a very short olecranon process, which is not known for other Late Miocene hominids. The femur is short and robust with a large head, and the knee joint indicates mobility in several planes. The hand is long but narrow, and the foot is comparatively short, though in both the hand and foot the digits are long and curved. The carpals and tarsals are primitive hominoid-like in being transversely gracile compared to their length.

*Oreopithecus* combines primitive and derived hominoid characters that ironically make it extremely difficult to place phylogenetically, despite its relatively complete preservation. Harrison and Rook (1997) consider *Oreopithecus* to be a stem hominid closely related to *Dryopithecus*. Moyà-Sola et al. (1997, 1999) interpret both *Dryopithecus* and *Oreopithecus* to be stem pongines, and they have also concluded that *Oreopithecus* was an arboreal biped with a well-developed precision grip. However, these conclusions are based in part on an erroneous reconstruction of the hand of *Oreopithecus* (Susman 2004 and personal observations) and a very unlikely reconstruction of the foot (Köhler and Moyà-Solà 1997 and personal observations). Rook et al. (1999) interpret CT scans of the innominate of *Oreopithecus* to imply a remodeling of bone consistent with bipedalism, but alternative interpretations are in my view more likely (Wunderlich et al. 1999). Overall, the overwhelming signal from the postcranium of *Oreopithecus* is of a suspensory arboreal adaptation. The long, curved phalanges are
unambiguous indicators of suspension and incompatible with either bipedalism or a precision grip.

Though some have interpreted aspects of the cranial morphology of *Oreopithecus* to have resulted from neoteny leading to a superficially primitive morphology (Moya-Sola et al. 1997; Alba et al. 2001), it is very difficult to identify heterochrony in fossil taxa (Rice 1997), and the much more straightforward interpretation is that *Oreopithecus* does in fact retain a number of primitive characters not found in other Late Miocene or extant hominids (Harrison 1986; Harrison and Rook 1997; Begun 2002). These include a short, gracile premaxilla, large incisive foramen, low position of the zygomatic root, small brain, a number of features of the basicranium and several postcranial characters (gracile phalanges, transversely small carpals, short, relatively gracile tarsal, etc.). It is very unlikely that a single growth process resulting from selection for bipedalism and an omnivorous diet, as suggested by Alba et al. (2001), would have produced such a diversity of consistently primitive characters throughout the skeleton.

The extraordinary morphology of the cranium and dentition of *Oreopithecus* are probably related to a specialized folivorous adaptation. *Oreopithecus* molars have the highest shearing quotients of any hominoid, which is consistent with a high-fiber diet (Kay and Ungar 1997). The exceptionally developed chewing muscles of *Oreopithecus*, its robust mandibles, and even the small size of its brain are all consistent with a folivorous diet requiring high-bite forces but relatively little planning or “extractive foraging” (Begun and Kordos 2004).

### 4.6 Late Miocene hominid extinctions and dispersals

Hominids first appear and quickly radiate in the Middle Miocene of Eurasia, but between about 10 and 9 Ma they begin to disappear. The view presented here is that the hominids from western Eurasia are hominines, and those in the east are pongines (Begun 2004). Descendants of each subfamily eventually disperse south of the Tropic of Cancer as other taxa become extinct in Eurasia (Begun et al. 1997; Begun 2001, 2004). This view has been supported by genetic evidence (Stewart and Disotell 1998) and criticized based on differing interpretations of the fossil record. For example, it has been noted that Africa is a more likely place for the origin of the Hominidae and the Homininae, but because it is poorly sampled, especially in the Late Miocene, the fossils that would support this interpretation remain to be discovered. In fact, many Late Miocene localities are known from Africa, a number with paleoecological indications of forested settings (Begun 2001, 2004), yet no hominids have ever been identified in Africa dating between
Kenyanpithecus and Sahelanthropus. A few specimens from this time period are known, such as Samburupithecus, but as noted this taxon retains many primitive dental and maxillary characters (Begun 2001). Isolated teeth from Ngorrora have been described as having affinities primarily with the Proconsuloidea or Middle Miocene East African hominoids (e.g., Equatorius) (Hill and Ward 1988; Begun 2001; Hill et al. 2002). Pickford and Senut (2005) have recently described teeth from Ngorrora and Lukeino have been described as chimpanzee and gorilla-like, but in my view the older teeth cannot be distinguished from others with affinities to the Proconsuloidea, and the younger teeth are probably from Orrorin, known from the same locality (Lukeino). An African origin of these clades also fails to explain the pattern in Eurasia that includes hominids with African great ape morphology in the west and Asian great ape morphology in the east (Begun 2004).

Hominids appear to have moved south from Eurasia in response to global climate changes that produced more seasonal conditions in Eurasia toward the end of the Miocene (Quade, et al. 1989; Leakey et al. 1996; Cerling et al. 1997; Begun 2001, 2004). Much evidence exists for climate change throughout much of Eurasia in the Late Miocene, which led to the development of more seasonal conditions. This culminates in the Messinian Salinity Crisis that led to the desiccation of the Mediterranean basin at the end of the Miocene (Hsü et al. 1973; Clauzon et al. 1996; Krijgsman et al. 1999). Other consequences include the development of Asian Monsoons, desertification in North Africa, the early phases of Neogene polar ice cap expansion and the expansion of North American grasslands (Garcés et al. 1997; Hoorn et al. 2000; Zhisheng et al. 2001; Griffin 2002; Guo et al. 2002; Janis et al. 2002; Liu and Yin 2002; Wilson et al. 2002). In both Europe and Asia, subtropical forests retreat and are increasing replaced by more open country grasslands and steppes (Bernor et al. 1979; Bernor 1983; Fortelius et al. 1996; Cerling et al. 1997; Bonis et al. 1999; Magyar et al. 1999; Solounias et al. 1999; Fortelius and Hokkanen 2001). In some places, forests persisted and elsewhere more severe changes occurred, creating a number of refugia, some of which continued to host hominids well into the period of climatic deterioration. This is the case for the Oreopithecus localities of Tuscany and Sardinia (Harrison and Rook 1997). Other well-known localities, such as Dorn-Dürkheim in Germany, retain a strongly forested character, though they lack hominoids (Franzen 1997; Franzen and Storch 1999).

There is a gradient of extinctions of forest forms from West to East corresponding to the gradient of appearance of more open country faunas from east to west (Bernor et al. 1979; Fortelius et al. 1996; Begun 2001, 2004). Between about 12 and 10 Ma Dryopithecus disappears from localities in Europe, becoming
very rare by 9.5 Ma in Spain and Germany. This wave of extinctions ends coincident with an important faunal event in Western Europe known as the mid-vallesian crisis, when a major turnover of terrestrial faunas leads to the widespread extinction of local taxa generally attributed to the development of more open conditions (Moya-Sola and Agustí 1990; Fortelius et al. 1996). The youngest specimens possibly attributable to *Dryopithecus* are the most easterly, currently assigned to *Udabnopithecus* from the 8- to 8.5-Ma locality of Udabno in Georgia (Gabunia et al. 2001).

In the eastern Mediterranean, hominids persist to the end of this time. *Ouranopithecus* in Greece is mainly known from the end of the hominid presence in Europe, and may be a terminal taxon of the *Dryopithecus* clade (Begun and Kordos 1997). In Anatolia at the eastern edge of the faunal province that includes Greece and the eastern Mediterranean (the Greco-Iranian province [Bonis et al. 1999]), a very large hominid resembling *Ouranopithecus* may be as young as 7–8 Ma in age (Sevim et al. 2001). At this time, forest taxa are increasingly replaced by more open country forms. This is true of virtually all mammalian orders. Among the primates, hominoids decline and cercopithecoids are on the increase (Andrews et al. 1996). Grazing ungulates and grassland or dry ecology adapted micromammals also become more common (Fortelius et al. 1996; Agustí et al. 1999; Bonis et al. 1999; Solounias et al. 1999).

The dispersal of Late Miocene faunas between Eurasia and Africa is complex and includes both open and more forest adapted taxa. Among the more open country taxa, horses disperse from North America to the Old World, and modern bovids and giraffids appear to have dispersed from Europe to Africa (Dawson 1999; Made 1999; Solounias et al. 1999; Agustí et al. 2001). Among the more close setting mammals, hippos move from Africa to Europe, and pigs of varying ecological preferences move from Asia to Europe and Africa (Fortelius et al. 1996; Made 1999). Small carnivores (mustelids, felids, and viverrids), larger carnivores (ursids, hyaenids) porcupines, rabbits, and chalicotheres, most of which also prefer more closed settings, also disperse from Eurasia to Africa (Leakey et al. 1996; Ginsburg 1999; Heissig 1999; Made 1999; Winkler 2002).

Many of these dispersals involved forest or wetter ecology taxa (hippos, some suids, primates, carnivores, rodents, and chalicotheres), which is consistent with the evidence of climate change at that time. Taxa disperse south into Africa as conditions continue to deteriorate leading to the Messinian crisis, among them probably the ancestors of the African apes and humans. This scenario has hominid ancestors leave Africa in the Early Miocene and return as hominines in the Late Miocene, but this is precisely what seems to have occurred in several mammalian lineages, including those represented by Late Miocene African species of *Orycteropus* (aardvark), several small carnivores, the hippo *Hexaprotodon*,
and possibly the proboscideans *Anancus, Deinotherium,* and *Choerolophodon* (Leakey et al. 1996; Ginsburg 1999; Heissig 1999; Made1999; Boisserie et al. 2003; Werdelin 2003; Begun and Nargolwalla 2004).

### 4.7 Summary and conclusions

The Miocene epoch witnesses several adaptive radiations of hominoids and hominoid-like primates. It was indeed the golden age of the Hominoidea. Many catarrhines appear in East Africa in the Early Miocene, some of which are surely related to living hominoids. A few of the basic attributes of the Hominoidea appear at this time, including the absence of a tail, somewhat extended life history, and a hylobatid level of encephalization, and hints of powerful hand and foot grips and a propensity for more vertical climbing. Among the diversity of Early Miocene Hominidea, a group emerged that may have had an adaptation to a diet dependent on more embedded resources, leading to a dispersal into Eurasia. Once there, hominoids flourish and expand, splitting into eastern and western clades that lead to extant hominids, and an early southern clade that becomes extinct. Early in the Late Miocene, the hominin radiation in Eurasia began to dwindle, with the earliest extinctions occurring in the west, and progressing eastward. Hominids and many other mammals experienced extinction events at this time, and many clades of Eurasian mammals also dispersed south, probably as a result of major global climatic events. Western Eurasian hominids dispersed into Africa leading the evolution of the African apes and humans, and eastern hominids dispersed into Southeast Asia, leading to the appearance of the *Pongo* clade. Shortly after their return to Africa, hominines diverged into their respective clades, probably relatively quickly. Gorillas remain the most conservative in many respects, though they achieve some of the largest body masses in any primate and specialize in their ability to exploit high-fiber keystone resources. Chimpanzees and humans diverged possibly within a million years of the emergence of the gorilla clade, the chimp clade remaining relatively conservative and the human clade experiencing much more rapid and dramatic evolutionary changes. Human ancestors retain the imprint of their Eurasia and African ape ancestors, and were very probably similar to extant African apes, particularly chimpanzees. That is, the fossil record of hominoid evolution suggests that humans evolved from a knuckle-walking, forest-dwelling soft fruit frugivore/omnivore. The details of the evolutionary events leading to the origin of the individual lineages of the Homininae remain to be worked out, a process hampered in part by a poor fossil record that, for example, includes almost no fossil relative of gorillas or chimpanzees (but see McBrearty and Jablonski 2005).
References


Begun DR (1992c) Dryopithecus crusafonti sp. nov., a new Miocene hominid species from
Begun DR (2004) Sivapithecus is east and Dryopithecus is west, and never the twain shall meet. Anthropol Sci ON LINE ISSN: 1348–8570
Begun DR, Ward CV (2005) Comment on “Pierolapithecus catalaunicus, a new Middle Miocene great ape from Spain.” Science 308: 203c
Begun DR (2005) Sivapithecus is east and Dryopithecus is west, and never the twain shall meet: Anthropol Sci, vol 113, p 53–64


Kordos L, Begun DR (1998) Encephalization and endocranial morphology in Dryopithecus brancoï. Implications for brain evolution in...
Lartet E (1856) Note sur un grand singe fossile qui se rattache au groupe des singes superieurs. CR Acad Sci 43: 219–223
Lydekker R (1879) Further notices of Siwalik Mammalia. Rec Geol Surv India XII: 33–52
McCrossin ML, Benefit BR (1997) On the relationships and adaptations of Kenyapithecus, a large-bodied hominoid from the middle Miocene of eastern Africa. In: Begun DR, Ward CV, Rose MD (eds) Function, phylogeny and


Pilgrim G (1910) Notices of new mammalian genera and species from the Tertiaries of India. Rec Geol Surv India 40: 63–71

Pilgrim G (1915) New Siwalik primates and their bearing on the question of the evolution of man and the anthropoidea. Rec Geol Surv India 45: 1–74


Rice SH (1997) The analysis of ontogenetic trajectories: When a change in size or shape is not heterochrony. Proc Natl Acad Sci 94: 907–912


Ruff CB, Walker A, Teaford MF (1989) Body mass, sexual dimorphism and femoral proportions of *Proconsul* from Rusinga and


5 The Biotic Environments of the Late Miocene Hominids

Jordi Agustí

Abstract

The habitat of the Middle and Late Miocene hominoids from western Europe, like *Dryopithecus*, was characterized by the prevalence of subtropical conditions. As a consequence, those environments were mainly dominated by fruit eaters and browsers, including a large variety of suids, cervids, rhinos, chalicotheres, and proboscideans. In contrast, in large parts of Eurasia, from eastern Europe (Greek-Iranian province) and northern Africa to China, the Middle Miocene Climatic Crisis led to the development of a xerophylous woodland, dominated by bovids, giraffids, and pursuit carnivores. At first, the worldwide dispersal of the hipparionine horses changed this scenario very little. However at 9.6 Ma, a significant event, the Vallesian Crisis, led to the extinction of most of the fruit eaters that had prevailed in the Middle Miocene European faunas. Hominoids persisted for a time in the Tusco-Sardinian Island and in the low latitudes of southwestern Asia. The worldwide spread of grasses between 8 and 7 Ma led to the final extinction of those hominoids. Hominoid evolution continued in eastern and southeastern Africa, in a habitat that strongly resembles that of the Greek-Iranian province.

5.1 Introduction

The Late Miocene, that is, the timespan between 11.6 and 5 Ma, is a crucial period in understanding the configuration of our present world. This is the time when the persisting laurophyllous evergreen woodlands (also called *laurisilvas*), which had spread over large parts of the Old World during most of the Miocene, were replaced by drier and more seasonal ecosystems, including savannas and steppes. It has been argued that the spread of this kind of environment may have played a key role in hominid evolution, by enhancing the appearance of new locomotor innovations, such as bipedalism (Lovejoy 1980; Coppens 1983). However, this change toward increasing dryness and seasonality was not a sudden one. Instead, this general trend appears to have been punctuated by a number of faunal and environmental events, which predated the onset of glacial–interglacial dynamics.
in the Northern Hemisphere. Significant physiographic and tectonic events appear also associated with this biotic turnover, such as the uplift of the Himalayas and the Tibetan Plateau, the final closure of the Atlantic–Indian seaway throughout the Mediterranean, and the desiccation of the Parathetys and Mediterranean seas. These paleogeographic events severely affected the continent–ocean interplay system, leading to changes in the overall circulation pattern, enhancement of the monsoonal dynamics, and the development of the first Arctic glacials.

5.2 The Middle Miocene environmental background: the world of Dryopithecus

Despite the climatic crisis that affected the Early Miocene terrestrial ecosystems, the last part of the Middle Miocene was characterized by very favorable biotic conditions in parts of western Eurasia, specially in western and central Europe (Agustí and Antón 2002; Fortelius et al. 2003). The faunal associations from these regions were characterized by high mammalian diversity levels, with a high amount of large-to medium-sized browsers, together with several medium-sized carnivores. At the top of the large browser guild were the proboscideans, represented by the large gomphotheres of the genus Tetralophodon and the deinotheres of the genus Deinotherium. Tetralophodon was a late immigrant in the Middle Miocene terrestrial ecosystems, and replaced the once worldwide gomphotheres of the genus Gomphotherium. Tetralophodon was a large proboscidean which had more hypsodont teeth than the earlier gomphotheres. Moreover, its skull, although still bearing four tusks, was shorter and more elephant-like, with a pair of long, straight tusks in the maxilla and a small pair of tusks at the end of the mandible. In contrast, Deinotherium represents a completely different kind of proboscidean. Members of this group had only two strong tusks, not placed in the upper jaw but at the end of the mandible. Moreover, this lower pair of tusks was recurved downward. The molars were very simple, formed of two cutting ridges, interpreted as an adaptation to browsing leaves and tough vegetation.

Another main component of the large browser community was the chalicotheres, a group of bizarre perissodactyles distantly related to horses that prolonged the trend observed in other ungulates to enlarge the forelimbs relative to the hindlimbs. This trend reached an extreme in some Miocene representatives, such as Chalicotherium, which developed gorilla-like limb proportions, with very long forelimbs and short hindlimbs. The chalicotheres bore claws at the end of their arms, instead of the typical hoofs of the odd-toed ungulates. These enabled them to grasp small branches and leaves with the help of their forefeet. These
large perissodactyls were probably capable of standing on their hindlimbs and using their forelimbs as “hands” to reach the higher vegetational levels.

However, the most diversified group of large browsers were the rhinoceroses, represented by several different forms, such as *Brachypotherium*, *Hoploaceratherium*, *Alicornops*, or *Lartetotherium*. *Brachypotherium* was a large teleoceratine with hypsodont teeth, short legs, and hippolike body proportions, which probably had a semiaquatic lifestyle. Another group were the aceratherine rhinos, represented by the genera *Hoploaceratherium* and *Alicornops*. *Hoploaceratherium tetrادactylum* was a medium-sized acerather with long limbs and slender body proportions. As in most aceratherines, the characteristic horns were highly reduced (only a small one was present). In contrast, they displayed very big lower incisors, which were larger in the males. *Alicornops simorrense* was a small acerather with short, tridactyl legs and strongly curved lower incisors (although, as in *Hoploaceratherium*, a small horn was present). The rhinocerotines or “modern horned rhinos” were represented by *Lartetotherium* and “*Dicerorhinus* steinheimsen*s*. *Lartetotherium sansaniense* was a cursorial rhino that had a unique, long horn. According to its rather brachydont teeth, its diet must have contained a higher quantity of soft plants and a lower proportion of wooden parts of shrubs than in the case of the aceratherine rhinos (Heissig 1989).

Represented only by the genus, *Anchitherium*, the equids were also significant elements of the Miocene ecosystems. *Anchitherium* was a member of the group of North American horses that experienced a significant evolutionary radiation during the Oligocene. This medium-sized anchitherine, about 1 m or less at the withers, crossed the Bering area in the Early Miocene and rapidly dispersed over the whole of Eurasia, from China to Spain. They bore low-crowned (brachydont), lophodont teeth adapted to browsing soft leaves. Their limbs were also adapted to locomotion on soft substrates, still retaining two lateral, fully functional toes.

The undergrowth of the Middle Miocene laurisilva was populated by a wide array of suids, which included hyotherines (*Hyotherium*), peccary-like suids (*Taucanamo, Albanohyus*), tetraconodontines (*Conohyus* and its offshoot *Parachleuastochoeramus*), and listriodontines (*Listriodon*). *Taucanamo* was a small peccary-like pig (about 12 kg), which developed lophodont teeth and large, elongated premolars. Its dental morphology, with high cusps and variable lophodonty, resembles that of some cercopithecoid primates and tragulids, being probably related to a browsing regime in a forest biome. *Albanohyus* was a small peccary-like suid also found at Fort Ternan that resembled *Taucanamo* but had smaller, shorter premolars. *Listriodon* was a fully lophodont, browsing listriodontine, which dispersed over the whole of Eurasia, from China to the Iberian peninsula throughout eastern and central Europe (Made 1996). Analysis
of microwear in *Listriodon* has shown a rather uniform diet with a smaller minerogenic component, which indicates a variation from the typical rooting behavior of generalized suids and a specialization in the browsing of vegetation. Lengthening of distal limb segments might indicate that these listriodontines preferred more open habitats. Together with the listriodontines, a new subfamily of suids, the tetraconodontines, became the dominant suiforms in the circum-Mediterranean area (including five genera: *Conohyus*, *Parachleuastochoerus*, *Sivachoerus*, *Notochoerus*, and the African *Nyanzachoerus*). The tetraconodonts bore thick-enameled cheek teeth and conical premolars with hyena-like wear, which probably indicates a diet based on hard food items such as seeds. A trend toward reducing size is present in this group during the Middle Miocene, from the 70 kg of the medium-sized *Conohyus* to the 40 kg of the small *Parachleuastochoerus*. Later on, at the end of the Middle Miocene, the suid diversity increased again with the appearance of the first representatives of modern suids (*Propotamochoerus*). *Propotamochoerus* was a large suid (about 120 kg), which probably evolved in Southern Asia from a hyotherine pig during the Middle Miocene and subsequently extended its range westward into southwestern Asia and Europe. It is the first recognizable member of all modern swines. The molars of this group show a trend toward the proliferation of several minor cusps, concomitant with loss of cusp identity. This peculiar dental evolution resembles some bears and indicates a further adaptation to omnivory.

Other significant members of the medium-sized browser community were the deer or cervids, represented by a number of genera such as *Heteroprox*, *Dicrocerus*, and *Euprox*. According to their limb proportions and low-crowned (brachydont) dentitions, most of these archaic deer were semiaquatic browsers that lived in closed forests in humid conditions (Köhler 1993). They still displayed rather simple, two-pronged antlers, although in some cases like *Euprox* there was a differentiation between a principal, posterior prong, and a secondary, smaller anterior prong. Some of them, like *Dicrocerus* and *Euprox*, showed a burrlike area, indicating for the first time the border between the deciduous and permanent segments of the antlers. Besides cervids, other related taxa, such as the moschid *Micromeryx* and the tragulid *Dorcatheirium*, testify to the persistence of very humid conditions in western and central Europe. *Micromeryx* was a very small and slender moschid of less than 5 kg that, as in the case of the living moschids, lacked horns but displayed very prominent canines (in the males). *Micromeryx* probably foraged in the lower vegetation of the closed forest, again living on soft plants and fruits, larvae, and carrion (Köhler 1993). It was a very successful moschid that spread over a wide area covering western and eastern Europe and that persisted until the Late Miocene (Early Turolian). Among these small browsers, other significant elements were the tragulids of the genus
Dorcatherium. The tragulids or chevrotains are small ruminants that today live close to the water in the closed forests of tropical Africa and East Asia. Like the moschids, they lack any kind of cranial appendages, having in their turn a pair of long canines which are longer in the males. Moreover, they retain a very primitive limb structure, with four well-developed toes on each foot (the two central metapodials are not yet fused in the cannon bone). The recent relatives of Dorcatherium, the chevrotains Hyemoschus aquaticus and Tragulus meminna, live in dense tropical forests close to rivers or water courses, which act as a refuge in the case of sudden attack by predators. Dorcatherium was nearly identical in all aspects to the recent Hyemoschus and probably developed a similar lifestyle.

In contrast, the bovids were poorly diversified at this time, mainly represented by Eotragus. The first bovids such as Eotragus probably originated in Asia, to the south of the Alpine belt. Later, in the Early Miocene, they colonized Europe and Africa simultaneously, their presence in this latter continent having been reported from Gebel Zelten (Libya) and also from Maboko and Ombo (East Africa). Eotragus was a small ruminant, the size of the living dik-dik or dwarf antelopes of the tribe Neotragini. The horn cores were short and conical, placed directly over the orbits. The teeth were brachydont, indicating a diet based on soft plants, fruits, larvae, insects, or even carrion. Their limb proportions were primitive, close to those of a cervid. They probably occupied a closed, wooded habitat where they probably ducked under the undergrowth (Köhler 1993).

At the opposite extreme, the small herbivore guild was extensively represented by a variety of rodents, including several species of hamsters (cricetids), dormice (glirids), eomyids, squirrels (sciurids), and beavers (castorids). The laur-ophyllous forests at this time were populated by a variety of flying rodents, such as flying squirrels of different sizes (from the large Albanensia to the small Blackia) and, probably, the eomyids Eomyops and Keramidomys. But other groups of small rodents, such as some glirids of the genus Glirulus, also developed “flying” forms with a patagium. This is very unusual among dormice and provides a strong indication of the persistence of closed forests at that time in western and central Europe. Another indication of the persistence of humid conditions in this part of the Old World is the frequent discovery of beavers in these faunas. One of them, Chalicomys, strongly resembled in size and morphology of the recent Castor fiber and, like the living form, was probably highly dependent on permanent rivers. A second smaller beaver, Trogontherium, is widely present at this time and was probably associated with more unstable environments.

Among the carnivores, the large predator guild was represented by amphicyonids and nimravids. The amphicyonids, or “bear-dogs,” resembled canids in their dentition but developed ursidlike characteristics such as very large size and robust canines (with double cutting edges in the case of the upper ones). The
heavy carnassials with horizontal abrasion were probably used for bone-crushing, an interpretation supported by the high sagittal crests in the skull, which housed a powerful musculature probably used for breaking bones. All these adaptations suggest that the amphicyons were probably occasional scavengers, at a time when truly scavenging specialists were still absent. However, despite these scavenging adaptations, the body plan of the amphicyonids indicates active hunting. They were probably active and agile predators practicing the solitary stalk-pounce hunting mode of recent felids and bears. Their long tails were probably used to balance the pounce, as in the case of the modern felids (Viranta 1996). The Middle Miocene records the first appearance of one of the most successful amphicyonid species in the Miocene, *Amphicyon major* (starting from Neudorf-Spalte and several localities in Spain, France, Germany, Czechia, Turkey). This was a large form, attaining the size of a lion. *Amphicyon major* had a long skull, with an elongated snout and relatively long and massive canines. The general limb proportions were similar to those of a bear, with short metapodials. Besides bear-like forms, like *Amphicyon major*, the amphicyonids also produced at this time hypercarnivorous forms such as *Thaumastocyon* and *Agnotherium*. Both were medium- to large-sized amphicyonids with a cursorial body plan that probably indicates pursuit hunting habits.

However, the typical hypercarnivorous predators of this time were the nimravids, a family of carnivores that developed felidlike, sabertooth adaptations. At the end of Middle Miocene, this family was represented by *Sansanosmilus jourdani*, a large species which reached 80 kg. The true felids were represented at that time by smaller species, all included in the genus *Pseudaelurus*: *P. turnauensis*, *P. lorteti*, and *P. quadridentatus*. The true ursids were represented at this time by two mesocarnivore forms, *Hemicyon* and *Plithocyon*, as well as a typical omnivore form, *Ursavus*. This last genus of small ursids was represented by *U. primaevus* of about 90 kg. The hyenids were represented by slender forms, such as *Protictitherium*, *Plioviverrops*, and *Thalassictis*. Different from the living members of this family, these primitive hyenids were more civetlike generalized carnivores than true scavengers. *Protictitherium* was an archaic insectivore/omnivore form, with very generalized dentition displaying a full set of premolars and molars. The postcranial skeleton suggests a semiarboreal existence and a diet consisting of small mammals, birds, and insects, far removed from the lifestyle of the modern hyenas. *Plioviverrops* was a mongoose-like insectivore/omnivore carnivore, which shows a progressive adaptation to insectivory, as indicated by the reduction of the sectorial portion of the dentition and the increase in the number of high, puncture-crushing cusps on the cheek teeth. Its skeleton was apparently more adapted to a terrestrial lifestyle than *Protictitherium*. *Thalassictis*, represented by the species *T. montadai* and *T. robusta*, was the first member of a hyenid lineage
characterized by wolflike meat- and bone-eating habits. They still retained an un-
specialized dentition, in some way similar to that of canids, although with a major emphasis on bone eating. The postcranial skeleton indicates terrestrial locomo-
tion but without special adaptations to cursoriality. Unlike Protictitherium and Plioviverrops, it indicates an adaptation to a more open woodland environment (Wenderlin and Solounias 1991, 1996).

Finally, at the very end of the Middle Miocene, a number of elements of probable African origin entered Eurasia again. This was the case, for instance, for the giraffids of the genus Palaeotragus. Palaeotragus was a relatively small, slender giraffid of about 250 kg bearing a pair of parallel ossicones standing upright over the orbits. They had long legs and limb proportions resembling those of the living okapi. The structure of the foot indicates that they were probably open country runners and, perhaps, good jumpers (Köhler 1993). They probably ate soft plants, mainly leaves, which they took hold of by grasping them with their long tongues. Besides Palaeotragus, other immigrants at the end of the Middle Miocene were the hominoid dryopithecines of the genera Pierolapithecus and Dryopithecus, which joined the previously existing pliopithecid anthropoids of the genus Pliopithecus. The coexistence of dryopithecids and pliopithecids is probably explained by the different dietary habits. While Pliopithecus and other pliopithecids were folivorous primates that settled on sclerophyllous woodlands, the dryopithecids, such as Dryopithecus, were frugivorous forms that lived in the canopy of the evergreen laurophyllous forests. In central Europe, the thin-enamed dryopithecids replaced the last thick-enamed hominoids of the genus Gripho-

5.3 The Greek-Iranian province

While, as we have seen, the Middle Miocene polar cooling and East Antarctic Ice growth did not imply a significant decrease in diversity in the evergreen woodland ecosystems of western and central Europe, its effects were much more severe in middle- to low-latitude terrestrial environments. There was a climatic trend toward cooler winters and decreased summer rainfall. Seasonal, summer-drought-adapted sclerophyllous vegetation progressively evolved and spread geographically during the Miocene, replacing the laurophyllous evergreen forests which were adapted to moist, subtropical, and tropical conditions with temperate winters and abundant summer rainfalls (Axelrod 1975). These effects are clearly seen in a wide area to the south of the Paratethys realm, extending from eastern Europe to western Asia. According to Bernor (1984), this region, known as the Greek-Iranian or SubParatethyan province, acted as a woodland environmental
“hub” for a corridor of open habitats which extended from western North Africa eastward across Arabia into Afghanistan, northwest into the eastern Mediterranean area and northeast into North China. The Greek-Iranian province records the first evidence of open woodlands through which a number of derived, open-country large-mammals, such hyenids, thick-enamelled hominoids, bovids and giraffids, diversified and dispersed into East Africa. The mammal composition of the Greek-Iranian province was very different from that of the more wooded environments that persisted in most of western and central Europe and approached in some ways that of the recent African savannas. This is why it has often been regarded as a “savanna-mosaic” (Bernor 1984) or “protosavanna” (Harris 1993) chronofauna. But in fact, as demonstrated by several analyses, this eastern biome was closer to an open sclerophyllous woodland than to the extensive grasslands present today in parts of Africa. Nevertheless, the peculiar biotope that developed in the Greek-Iranian province acted as the background from which the African savannas evolved during the Plio-Pleistocene (Solounias et al. 1999). They also include a number of genera common to the open-country chronofauna, which dominated the Late Miocene Old World. This evolution has been documented in East Africa, where a similar ecosystem of seasonally adapted, sclerophyllous woodland, with terrestrial hominoids (*Kenyapithecus*), was present in Fort Ternan (Kenya) as early as 14 Ma, in association with the first grasslands in eastern Africa (Dugas and Retallack 1993).

At the taxonomic level, this habitat change in the western European low latitudes involved the rapid adaptive radiation of woodland ruminants (bovids and giraffids). Thus, although the small *Eotragus* persisted and is found to have been widespread throughout Europe, western Asia, and Africa, a new group of larger bovids, the boselaphines, spread at this time, becoming the most successful elements of this family during most of the Miocene. The boselaphines, today represented by the nilgai (*Boselaphus tragocamelus*) from India, began a successful evolutionary radiation in the Middle Miocene, which led to a high generic diversity. *Miotragocerus*, the first boselaphine to appear in Europe, is found from Byelometcheskaya in the Caucasus to Tarazona in central Spain. It was a medium-sized bovid of about 80 kg with strong horn cores that looked very different from those of *Eotragus*. Its teeth were still primitive but with some cementum. The limb bones and foot anatomy indicates that *Miotragocerus* lived in very humid habitats, where it probably fed on soft plants (Köhler 1993). At that time, a second boselaphine bovid in the Greek-Iranian province, *Austroportax*, displayed a quite different aspect. *Austroportax* was a large and surprisingly advanced bovid for its time. It weighed about 300 kg and was supported by short, heavy extremities resembling those of the living buffalos and other modern...
members of the tribe Bovini. Its foot morphology indicates that it lived in humid and wooded habitats.

A second group of successful bovids that spread at that time over southern Europe and Africa originated from a Middle Miocene Asian form called Caprotragoides. Following the Middle Miocene environmental changes, Caprotragoides spread over Europe and Northern Africa throughout the Greek-Iranian province, leading to Tethytragus and Gentrytragus, respectively. They were medium-sized bovids of about 30 kg with horn cores curved backward and slightly outward. The teeth morphology indicates a diet based on a great variety of plants. Most of their characters seem to be adaptations to dwelling in open country, but others indicate more wooded preferences. This mosaic of features indicates that Caprotragoides, Tethytragus, and Gentrytragus were probably eurytopic bovids with a high capability of invading very diverse biotopes.

A third group of advanced bovids that spread at this time were the hypsodontines. Hypsodontus was a medium to large (about 110 kg) slender and specialized long-legged bovid. It differed from the Boselaphini, Eotragus, or Tethytragus in its extremely hypsodont cheek teeth, indicating a diet based on grass and tough plants. It attained a broad Old World distribution in the Middle Miocene, from China to India, eastern Europe, and Africa. A second genus related to Hypsodontus, Turcocerus, was present in Turkey at the same time. Turcocerus was a very small bovid with slender, though massive, metapodials. It bore two short conical horns showing a clockwise torsion. The teeth were also very hypsodont and with cement, indicating a diet based on leaves, herbs, and grasses.

A fourth group of advanced bovids that spread in the Middle Miocene were the Antilopini, mainly represented by the gazelles. The first gazelles (Gazella) come from the Early Miocene beds of the Chinji zone of Siwaliks and from Majiwa in Kenya (Thomas 1984). According to these data, Gazella and other Antilopini could have originated in Africa or the Siwaliks from a form close to Homoiodorcas or a related Neotragini. Gazelles dispersed into Europe at this time from their possible Afro-Arabian origins, perhaps taking part in the same dispersal event as Giraffokeryx and the kubanochoers (see below).

Not only gazelles but also the giraffids experienced a wide adaptive radiation in Africa after their dispersal from Asia. One of these giraffids, Giraffokeryx, dispersed out of Africa and became widespread at this time, its remains having being found at several Middle Miocene localities of the Greek-Iranian province, such as Paçalar and Prebreza, as well as in the Bugti beds of Pakistan. Giraffokeryx displayed two pairs of rather short, unbranched ossicones. The anterior pair was situated in front of the orbits, while the second pair arose directly behind the orbits. A second giraffid, Georgiomeryx, has also been found in some Middle
Miocene localities of the Greek-Iranian province, like Chios in Greece, and Byelometscheskaya in the northern Caucasus. Georgiomeryx was closely related to Giraffokeryx but displayed a unique pair of flat, laterally extended ossicones over the orbits and a more archaic dentition with brachydont teeth (de Bonis et al. 1997a).

In contrast with this highly diversified bovid and giraffid fauna, the cervoid representation in the open woodland areas of the Greek-Iranian province was extremely poor, almost reduced to primitive moschoids of the genera Hispanomeryx and Micromeryx. Suids never attained the high-diversity levels observed in western and central Europe, the ubiquitous Listriodon splendens becoming the most common element. The listridontines evolved in a peculiar way in North Africa, leading to giant forms such as Kubanochoerus, which may have reached 800 kg in some cases. Kubanochoerus was found for the first time in the Caucasus (Byelometscheskaya) and probably derives from the African Libyochoerus (from the Early Miocene locality of Gebel Zelten). The most striking feature of these giant listriodontines was the presence in the males of an enormous horn above the orbits, which was probably used for intraspecific fighting and which indicates a unique case of territoriality in suids.

While the tetraconodonts, such as Conohyus, were dominant in western Europe, a group of archaic, small-sized suids, the sanitheres, persisted and succeeded in the Greek-Iranian province by developing selenodont cheek teeth. Their molarized premolars and molars bore wrinkled enamel formed of several cusps and ridges, well adapted to browse on sclerophyllous vegetation (de Bonis et al. 1997b). A second browsing pig was Schizochoerus, a small peccary-like suid related to Taucanamo that developed lophodont molars and short, broad premolars. This dentition resembles that of the contemporaneous advanced listriodontines and, as in that group, was probably well adapted to browsing in the vegetation of the sclerophyllous evergreen woodland, which covered most of the Greek-Iranian province at that time.

From the Asian side, members of the central Asian aceratherine hornless genus Chilotherium became the most common rhinos in the Greek-Iranian province. They were a group of grazing animals that occupied different niches and radiated into a number of (sub)genera such as Subchilotherium or Acerorhinus. Their legs were shorter than in any other aceratherine, mimicking those of the teleoceratines. A few of them were still clear browsers, like the brachydont Acerorhinus, while most of them were grass eaters (although certainly their diet included a number of nongraminean herbs). The shortening of the legs in this group can be explained by this grass-based diet. As aceratherines, they were hornless rhinos equipped with tusks-like incisors probably used in fighting. Accordingly, but in contrast with the living grass-eater rhinos, the head maintained
a horizontal position, so grazing was only possible after the shortening of the legs (Heissig 1989).

This time also records the first appearance of the hyenids (*Protictitherium*) in eastern Europe and western Asia. However, this does not mean in any way that the hyena-like scavenger niche was empty at that time, since a peculiar family of carnivores, the percocutids, occupied that place in the Greek-Iranian province. The percocutids seem to correspond to an early feloid radiation covering the “hyena guild,” at a time when the true hyenids (*Protictitherium*) had not yet developed the dental and locomotory adaptations to scavenging and the bone-cracking characteristic of the later members of the family. The first percocutids belong to the genus *Percocuta* and are found in late Middle Miocene localities of western (Sansan, La Grive) and eastern Europe (Çandir, Paçalar), where they tended to coexist with the small, arboreal primitive hyenids of the genera *Protictitherium* and *Plioviverrops*. The Middle Miocene *Percocuta* had not yet developed the bone-cracking adaptations which would be common in the Late Miocene members of the family (*Dinocrocuta*).

### 5.4 Southwestern Asia: the environment of *Sivapithecus*

In contrast with the Greek-Iranian province, conditions were very different to the south of this region. Actually, the environment in southwestern Asia seems to have remained much closer to that in western and central Europe. According to the rich mammalian record of the Siwalik sequence in the Potwar Plateau (northern Pakistan), warm tropical-subtropical forest zones persisted in this area during most of the Miocene. Therefore, the Chinji and Nagri faunas, equivalent to those of the late Middle and early Late (Vallesian) Miocene of Europe, are basically composed of a mixed assemblage of archaic carnivores, poorly diversified browsing ruminants, and woodland/bushland omnivores (Bernor 1984; Barry et al. 1985). A number of large- to medium-sized browsers are reminiscent of the Greek-Iranian province, such as the sivatherine giraffids of the genus *Giraffokeryx* (*G. punjabiensis*), the rhino *Chilotherium intermedium*, or the small suid *Schizochoerus gandakasensis*. Despite these common elements, the faunas from the Kamlija and Chinji beds are very different from those of the Greek-Iranian province (including the Middle Miocene sites of eastern Africa such as Fort Ternan), maintaining low-diversity levels of bovids and giraffids. Although the species are different, the Siwalik record includes taxa that are similar to those of the coeval western and central Europe faunas: proboscideans (*Deinotherium* sp.), rhinoceroses (*Brachypotherium permense*), chalicotherids
(Chalicotherium salinum), suids (Listridon pentapotamidae, Conohyus sindiensis, Propotamochoerus hysudricus, Hippopotamodon sp.), tragulids (several species of Dorcatherium and Dorcabune), carnivores (Agnotherium and other amphicyonids, nimravids like Sansanosmilus and Barbourofelis), cricetids (Democricetodon sp., Megacricetodon sp.), flying squirrels and shrews. As in the case of western Europe, this assemblage strongly points to a rather closed, forested environment. Particularly significant is the presence of hominoids with climbing locomotory adaptations of the genus Sivapithecus (Pilbeam et al. 1996).

A case that is peculiar to the Siwaliks, and is not shared with western or eastern Europe, is the persistence of the anthracotherids, a family of archaic artiodactyls distantly related to the hippos. The anthracotheres were large suiforms with selenodont molars well adapted to a browsing regime. However, unlike the living suids, they still retained five digits on the forefeet and four on the hindfeet (although the lateral ones were more reduced). This was a unique combination of ruminant-like, selenodont dentition, coexisting with a generalist, pig or hippolike shape, with relatively short, stout legs and still functional lateral digits (actually, a combination very close to that of the living hippos). In the Middle to Late Miocene of Siwaliks, the anthracotherids are represented by two species of different size, Microbunodon punjabiense and Hemimeryx sp.

The observed differences between the sclerophyllous woodland faunas of the Greek-Iranian province and those from the Siwaliks can probably be explained on the basis of the lower latitudinal location of the latter region and the influence of the rising Tibetan plateau on the development of the monsoonal climatic regime (Kutzbach et al. 1993). The Potwar Plateau is flanked to the north and the west by north–south lying Sulaiman and Kirtar ranges and Baluchistan and Sind. The uplift of these encircling mountain ranges would have trapped moist Indo-Pacific monsoons of the slopes facing the Siwaliks and nourished the less seasonal environments there (Bernor 1984).

5.5 The Hipparion dispersal event

Between 12 and 11 Ma, a drastic cold pulse led to a new growth of the Antarctic Ice Sheet and a global sea-level fall of about 140 m (Haq et al. 1987). The oceans dropped about 90 m below the present sea level and a number of land-bridges came again into existence, thus enabling faunal exchange between previously isolated terrestrial domains. As a consequence, a new corridor was reestablished between Asia and North America across what is now the Bering isthmus. The main result of the reopening of this land-bridge was the quick dispersal into Eurasia of the hipparionine horses of the genus Hipparion and their relatives.
The hipparionine horses arose in North America during the Middle Miocene and differed significantly from Anchitherium and other similar equids in the development of very high-crowned cheek teeth as a response to the more sclerophyllous, harder vegetation. Moreover, the tooth enamel became folded in several ridges, which were in their turn filled with dental cementum. The two persisting lateral toes in the hipparionine horses became more reduced than in Anchitherium, thus concentrating most of the body weight on the central toe. After the establishment of the Bering land-bridge, the hipparionine horses quickly invaded the whole of Eurasia, from China to Iberia, their presence having been reported from hundreds of fossiliferous localities. Existing data suggest that, after their entry into Eurasia, the hipparionine horses spread very quickly across Europe, their presence having been reported at 11.1 Ma both in the Vienna and the Vallès-Penedès Basins (Garcés et al. 1997). They probably colonized first the more northern latitudes of Asia and spread later to the south and east. The dispersal of the hipparionine horses appears therefore as an Old World event and defines the lower boundary of the Vallesian Mammal Stage, the continental equivalent of the early Late Miocene in Eurasia. Hipparion primigenium, the first hipparionine species to enter Europe, was a relatively large form standing about 1.5 m at the withers (the stature of a Burchell's zebra). Its slender axial skeleton suggests it was well adapted for leaping and springing rather than for sustained running and high speed (Bernor and Armour-Chelu 1999). This and other archaic hipparionine horses are included by some authors in the separate genus Hippotherium.

Although a single taxon event, the dispersal of Hipparion dragged on other immigrants from the open woodlands of central and western Asia into the laurophyllous forests of western Europe. This was the case for the first European leporids of the genus Alilepus, the sivatherine giraffids of the genus Decennatherium, and the saber-tooth felids of the genus Machairodus. Among lagomorphs, the leporids (the family that includes the living hares and rabbits) had a long evolutionary history in North America since the Eocene. However, it was not until the Early Vallesian that they settled in Europe, at that time still dominated by the pikas of the genus Prolagus. Another immigrant in this time, Decennatherium, was one of the first members of the sivatherines, a lineage of large, robust giraffids which differed from the more slender Palaeotragus in the possession of not two but four ossicones, a first pair over the orbits and a second larger pair at the rear of the skull. The sivatherines became the dominant giraffids of the Late Miocene terrestrial ecosystems and persisted in Africa until the Early Pleistocene, coexisting with the first hominids. As in the case of other primitive sivatherines, such as Bramatherium and Hidaspitherium, from the Siwalik Hills in Pakistan, Decennatherium probably had an enlarged anterior pair of
ossicones (or a unique fused anterior ossicone) and a less prominent posterior pair. Its limb bones were longer and more slender than the later members of this group.

Another typical Early Vallesian newcomer was Machairodus, a large saber-toothed cat that coexisted with the last nimravids of the genus Sansanosmilus. Apart from the large amphicyonids, all the other Middle Miocene hypercarnivorous predators were relatively small forms of less than 100 kg, but members of the genus Machairodus were large saber-toothed cats which could attain 220 kg (the size of a lion). As in the case of nimravids, the most characteristic feature of these predators was their long, laterally compressed and flattened upper canines, which greatly surpassed the size of the lower ones. The first machairodontine cats are recorded in the Middle Miocene of the Greek-Iranian province and persisted there until the Early Vallesian (Miomachairodus pseudailuroides from Yeni-Eskihisar and Eçme-Akçacoy in Turkey). Machairodus aphanistus is the most common Late Miocene species in Eurasia, ranging from the Iberian Peninsula to North America. Its limb anatomy was very different from that of the modern cats, with forelimbs longer and more robust than the posterior limbs, a feature which enabled them to grasp and immobilize their prey (Turner and Antón 1997). It seems that, at a first glance, the entry of Machairodus, a felid filling the large predator guild previously occupied by Sansanosmilus, would have had serious consequences for these nimravids, including their final extinction by competition. However, this was not the case and both Machairodus and Sansanosmilus coexisted for more than a million years without replacement of one by the other. A similar case was found in the Greek-Iranian province, where the large nimravid Barbourofelis coexisted with Miomachairodus in the Early Vallesian beds of the Sinap Formation in Turkey. Barbourofelis was larger than Sansanosmilus, the size of a lion, and in the Late Miocene attained a very broad distribution, from North America to eastern Europe.

The diversity also increased among the large browser perissodactyls such as rhinos and tapirs. From the rhino side, the most common form at that time was Aceratherium. This hornless aceratherine, close to the Middle Miocene Hoploaceratherium, was one of the most long-lasting genera of the Late Miocene, surviving until the Miocene–Pliocene boundary about 5 Ma. It was a medium-sized rhino with long limbs and a still functional fifth metacarpal. The cheek teeth were brachydont, indicating a browsing diet based on leaves and soft vegetation. Its limb proportions, close to those of the living tapirs, suggest a similar lifestyle (Heissig 1989). The males of Aceratherium incisivum bore a pair of strong tusks which enabled them to browse the dense vegetation of the Early Vallesian laurophyllous woodlands. As in other aceratherines, these tusks were much smaller in the females. In the Greek-Iranian province, other advanced rhinos of probable
African origin joined *Chilotherium* in the Late Miocene. This was the case of *Ceratotherium*, the genus that includes the living white rhino. The early representatives of *Ceratotherium* (*C. neumayri*) were only partly grass eaters, but a trend is observed in this group throughout the Miocene and the Pliocene to develop more open-country adaptations such as large body dimensions and slightly hypsodont cheek teeth.

Another group of perissodactyls that flourished at this time were the tapirs (*Tapirus priscus*), which reappeared in western Europe after their disappearance in the very Early Miocene. The anchitherine horses were represented by larger, more advanced species of *Anchitherium* which developed longer limbs and higher-crowned dentitions. These anchitherine horses persisted in central Europe until the latest Vallesian, although in some regions (Spain) they disappeared shortly after the entry of the first hipparionine horses. Among the smaller browsers, peculiar elements in Europe were the hyraxes, which are found at a number of Vallesian localities such as Can Llobateres (Vallès-Penedes, Spain), Melambes (Crete, Greece), and Eçme-Akçakoy (Turkey). Despite its small size and rabbit-like appearance, the hyraxes are archaic ungulates that today inhabit the rocky and steppe environments of central and southern Africa (although its range extends up to Lebanon). Their molars are brachydont and selenolophodont, strongly resembling those of some archaic perissodactyls. Although the living hyraxes are hare sized (about 50 cm in length), the Late Miocene European forms such as *Pliohyrax* reached large body dimensions, comparable to those of a tapir. The first hyraxes that settled in Europe in the Early Vallesian certainly had an African origin. Therefore, a limited exchange with northern Africa still existed in the Early Vallesian, although the possibility that hyraxes entered with the dryopithecids at the end of the Middle Miocene cannot be excluded.

Among the carnivores, the large amphicyonids of the species *Amphicyon major* persisted. The hypercarnivorous and cursorial amphicyonids *Thaumastocyon dirus* and *Agnotherium antiquus* persisted also in the Early Vallesian. *Agnotherium antiquus* is a poorly known species present in several localities from western and central Europe (Pedregueras, Eppelsheim, Rudabanya) and known also from northern Africa (Bled Douarah). It was similar to the better known Middle Miocene *Agnotherium grivensis* but smaller in size (160 kg). Among the bears, the small omnivorous ursids of the genus *Ursavus* diversified into the species *U. brevirhinus* and *U. primaevus*, while the mesocarnivore *Hemicyon* was represented by *H. goeriachensis* (of about 120 kg). But the most significant event in this group was the appearance of the first large Ursids of the genus *Indarctos*. *Indarctos vireti*, of about 175 kg, was the first member of a lineage of large mesocarnivore ursids, which were characteristic elements of the Late Miocene carnivore community.
In the Early Vallesian, the true hyenids were still represented by the civetlike Protictitherium, the mongoose-like Plioviverrops and the wolflike Thalassicis. At this time, Protictitherium also colonized northern Africa, being present in the Early Vallesian of Tunisia (P. punicum). The cursorial canidlike hyenid guild was enriched with new forms close to Thalassicis such as Ictitherium and Hyaenictitherium. Both may have been originated in the Greek-Iranian province or elsewhere in Asia and spread later into Europe. This time also records a high diversity of old viverrids (such as Semigenetta ripolli from Can Llobateres) and mustelids, which inherited the variety of forms present at the Middle Miocene: badgers (Sabadellictis), skunks (Promephitis, Mesomephitis), otters (Sivaonyx, Limnonyx, Lutra), volverines, and glutton-related forms (Trochictis, Circamustela, Marcetia, Plesiogulo). Some of these glutton-related forms were relatively large sized for a mustelid, reaching 50 kg in the case of Hadrichtis and Eomellivora.

Therefore, despite its significant zoogeographic importance, the spread of the hipparionine horses and their cohort of Asian immigrants was a quite limited event which did not result in a significant change in the structure of the previously existing western Old World mammalian communities. Newcomers like Machairodus, Alilepus, Hipparion, or Decennatherium joined the already highly diversified western European faunas without a clear and immediate replacement of the potential competitors which were there occupying similar guilds. The same situation is observed in the Potwar Plateau, where no significant mammal turnover is associated with the entry of Hipparion. Only the extinction of older species, notably suoids and cricetids, is recorded at the base of chron C5N (ca. 10.8–10.9 Ma), shortly before the first occurrence of this equid in the Siwaliks sequence (Barry et al. 1985; Pilbeam et al. 1996). The Early Vallesian faunas in Europe are thus characterized by the “peaceful” coexistence of a number of species which seem to have filled similar ecological guilds. This led to a sort of “climax” situation in the western European ecosystems, which reached levels of mammalian diversity unknown in any other Late Cenozoic epoch. With more than 60 mammal species, localities such as Can Ponsic and Can Llobateres 1 in Spain or Rudabanya in Hungary are good examples of these Early Vallesian “inflated” faunas. Despite the presence of new immigrants like Hipparion, Decennatherium, or Machairodus, the western European Vallesian ecosystems were composed of almost the same elements that populated the Middle Miocene subtropical forests, retaining a similar community structure. In this environmental context, the Eurasian hominoids reached an extraordinary diversity, which included the forest-adapted, suspensor Dryopithecus and Sivapithecus, the dry-adapted Ankarapithecus, and the robust, gorilla-like Graecopithecus.
The Vallesian Crisis

After the high-diversity levels attained in the Early Vallesian, an abrupt decline in the Vallesian mammalian faunas took place at about 9.6 Ma in what is known as the Vallesian Crisis (Agustí and Moyà-Sola 1990). The Vallesian Crisis was first recognized in the Vallès-Penedès Basin of Spain and involved the sudden disappearance of most of the humid elements that characterized the Middle Miocene and Early Vallesian faunas from western Europe. Among the large mammals, this crisis particularly affected several groups of perissodactyls such as the rhinoceroses *Lartetotherium sansaniense* and “Dicerorhinus” *steinheimensis* and the tapirs (only the small tapirs of the badly known *Tapiriscus pannonicus* persisted until the Early Turolian in central Europe; Franzen and Storch 1999). These losses were only partly compensated by the entry just before the onset of the Vallesian Crisis of *Dihoplus schleiermacheri*, a large species of browsing rhino which bore a pair of massive horns and was the largest rhino of its time. Among the artiodactyls, the high diversity attained by the suids in the Early Vallesian times suddenly dropped and several characteristic elements vanished. This was the case for the browsers *Listriodon* and *Schizochoerus* (which made a short-lived incursion in western Europe at the beginning of the Late Vallesian), as well as for the tetracodontines *Conohyus* and *Parachleuastochoerus*. In contrast, the “modern” suinae, such as *Propotamochoerus*, persisted and even enlarged their diversity with a new eastern immigrant *Microstonyx*. *Microstonyx* was a giant pig (about 300 kg), with a skull more than half a meter long. The Vallesian Crisis also involved the final decline of the Middle Miocene forest community of cervoids (the cervid *Amphiprox* and the moschid *Hispanomeryx*) and the spread of the boselaphine bovids like *Tragoportax*, which replaced their semiaquatic relatives of the genus *Protragocerus*. *Tragoportax* was a medium-sized bovid of about 80 kg with relatively long limbs, which suggests that it was a fast runner and a good jumper which lived in the open woodland. It possessed a short-faced skull with a long neurocranium and large backwardly curved horns. The teeth were high crowned and with cementum, resembling those of the living *Boselaphus* (Köhler 1993).

Among the rodents, the Vallesian Crisis involved the disappearance of most of the cricetids and glirids of Early or Middle Miocene origin (*Megacricetodon, Eumyarion, Bransatoglis, Myoglis, Paraglirulus, Eomuscardinus*), flying squirrels (*Albanensia, Miopetaurista*), and beavers (*Chalicomys, Euroxenomys*). However, other less diversified small mammal groups, such as lagomorphs and insectivores, remained almost unaffected by this crisis. In western and central Europe, this event coincided with the first dispersal of the murid rodents, the family that include the living mice and rats. After their entry into Europe, this group became
the dominant rodents in the Late Miocene communities and diversified into a number of genera: Progonomys, Occitanomys, Huerzelerimys, Parapodemus.

Another group which was severely affected by the Vallesian Crisis was the large carnivores of the families Nimravidae and Amphicyonidae. Among the amphicyonids, all the genera and species still existing in the Early Vallesian disappeared: Pseudarctos bavarius, Amphicyon major, Thaumastocyon dirus. Only some poorly known Amphicyon representatives persisted in the Late Vallesian and Early Turolian in some parts of central Europe. Among the ursids, the Vallesian Crisis had an ambivalent effect. While the slender cursorial forms of Early Miocene origin like Hemicyon vanished, the robust ursids of “modern” aspect persisted, represented by larger species. This was the case with Indarctos, represented by the species I. vireti and I. arctoides, as well as Ursavus represented by U. depereti, the largest species of the genus. In turn, the mustelids were severely affected by the Vallesian Crisis, which involved a significant decrease in the once highly diversified Vallesian fauna.

At the same time, a number of eastern immigrants appeared for the first time such as the large hyenids of the genus Adcrocuta and Hyaenictis. These genera represent two opposite trends in the evolution of Late Miocene hyenids. Hyaenictis was a cursorial meat and bone eater, which prolonged the trend initiated by Thalassictis toward increasing cursoriality. They show also a trend toward the reduction of the bone-crushing portion of their dentition, developing and extending at the same time the sectorial part, so that the posterior molars were reduced or lost. At the other end of the scale, Adcrocuta, at about 70 kg, was the first representative of the modern bone-cracker hyenids leading to the living Crocuta and Parahyaena. They were characterized by advanced adaptation to bone crushing, with enlarged bone-cracking premolars. Adcrocuta had short stocky limbs, indicating that it was not a cursorial form. Like Hyaenictis, it was of probable Asian origin.

Among the large predator guild, the nimravids finally came to an end with the Vallesian Crisis, after representing the “large cat” guild for millions of years. Their extinction was compensated with the entry of Paramachairodus, a new genus of machairodontine cats. Paramachairodus ayygja, the oldest species of this genus, was smaller and more slender than Machairodus aphanistus (about 44 kg), retaining the archaic anatomy inherited from its ancestor Pseudaelurus quadridens. With their robust forelimbs and slender hindlimbs, the members of this species were probably able to climb trees, carrying large prey as do living leopards. Their long muzzled skull was also superficially leopard-like, although, as a true machairodont, their upper canines were characteristically long and laterally flattened (Turner and Antón 1997).
Last but not least, the Vallesian Crisis led to an abrupt end of the hominoid experiment in Europe. Hominoids like *Dryopithecus*, *Ankarapithecus*, or *Graecopithecus* disappeared entirely from the fossil record, and only *Oreopithecus* in its island refuge and *Sivapithecus* in southwestern Asia survived this extinction event (see the next sections). *Dryopithecus* is still found in some early Late Vallesian localities dated at about 9.6 Ma (Can Llobateres 2, Viladecavalls) but disappeared from the fossil record shortly after. In the Greek-Iranian province, the robust *Graecopithecus* also disappeared at the beginning of the Late Vallesian. A similar case was that of the Turkish *Ankarapithecus*, its record ending again in the Late Vallesian. The extinction of these robust hominoids in this province coincides with the spread of the colobine monkeys of the genus *Mesopithecus*. This was not the case in western Europe, where the extinction of the slender dryopithecines did not involve its replacement by any other kind of primate species. Only the persistence until the latest Vallesian of the advanced folivorous pliopithecids of the genus *Anapithecus* can be quoted in this area, some hundred thousand years after the last *Dryopithecus*. In China, the pliopithecids survived even longer, until the latest Miocene, being represented by a large-sized form (*Laccopithecus*).

5.6.1 Causes of the Vallesian Crisis

What could have caused the set of extinctions and deep faunal restructuring which took place at 9.6 Ma during the Vallesian Crisis? Some evidence, such as the exit of several forest forms and the development of sigmodont teeth by some groups of rodents, would support the replacement at that time of the laurophyllous forests by grasslands. However, the spread of grasses over large extensions of Eurasia has been dated by Cerling and coworkers to between 8.3 and 7 Ma, and geochemical analyses carried out on teeth and soil nodules older than that timespan do not detect any sign of such an environmental change. Nevertheless, the fact that this major ecological restructuring of the western European mammal assemblages affected especially those taxa with tropical forest affinities, and the latitudinal character of these extinctions (a number of forest-adapted taxa survived until the Early Turolian in central Europe), strongly suggests its climatic forcing. This is supported by the Late Miocene oceanic evolution, which was a continuation of the processes started at the Middle–Late Miocene transition. Enhancement of the latitudinal thermal gradient resulted in the generation of new erosive oceanic surfaces (NH5) by intensification of the deep circulation. Further cooling resulted in new $\delta^{18}O$ positive shifts (i.e., Mi6 and Mi7; Miller et al. 1987, 1991). Changes in benthic and planktonic assemblages also indicate
colder climatic conditions and increasing isolation between low and middle latitudes. All these oceanic changes were nearly synchronous with some significant changes in low-latitude Old World terrestrial domains. In particular, there is a noticeable synchronism between the Mi7 isotopic shift at 9.3–9.6 Ma and the age of 9.6 Ma obtained for the Vallesian Crisis in the Vallès-Penedès stratigraphic sections. The Vallesian Crisis is also close to the beginning of the NH5 hiatus, one of the most important sets of deep oceanic discontinuities recognized in the Late Miocene (Keller and Barron 1983), which is dated between 9.0 and 9.5 Ma. NH5 has been also related to a period of cooling and major restructuring of the deep oceanic circulation and to the growth of ice sheets in western Antarctica (Keller and Barron 1983).

But how did these changes affect the composition of the terrestrial vegetation? A change to more open environments did not start until 8 Ma, the extension of grasslands taking place almost 2 Myr after the Vallesian Crisis. The pre-Vallesian floras in the region indicated the persistence of humid subtropical conditions, with abundance of broad-leaved mega-mesotherm elements such as Ailanthus, Caesalpinia, Cassia, Cinnamomum, Ficus, Sapindus, etc. (Sanz de Siria 1994). Indeed, the very rich mammal locality of Can Llobateres 1 yielded remains of some of these subtropical elements such as Sabal, Ficus, and others. However, we know that the laurophyllous subtropical woodland prevailing until the beginning of the Late Miocene in Europe was profoundly affected in some way. The answer arrived in 1994, when for the first time a well calibrated Late Vallesian flora was discovered in a section close to the city of Terrassa in the Vallès-Penedès Basin (Sanz de Siria 1997). This flora has been dated by paleomagnetism at somewhat more than 9 Ma and therefore records the kind of vegetation that was dominant just after the Vallesian Crisis (Agustí et al. 2003).

The flora from the Terrassa section includes 36 different taxa from the families Lauraceae, Ulmaceae, Hamamelidaceae, Juglandaceae, Myricaceae, Fagaceae, Betulaceae, Tiliaceae, Salicaceae, Ericaceae, Sapotaceae, Myrsinaceae, Celastraceae, Aquifoliaceae, Rhamnaceae, Sapindaceae, Aceraceae, Oleaceae, Poaceae, and Typhaceae. Close to 45% of this flora is composed of deciduous trees such as Acer, Alnus, Fraxinus, Carva, Juglans, Populus, Parrotia, Zelkova, Ulmus, or Tilia, that is, the elements that are now dominant in the temperate forests of the middle latitudes. In contrast, warm evergreen elements, represented by Myrsinie, Sapindus, Sapotacites, etc., decreased to 7%. Supporting the argument that this change was not related to the extension of grasslands was the persistence of a “hard core” of subtropical elements, represented by 33% of evergreen trees, able to endure a certain level of seasonality, and which persisted in Europe until the Early Pliocene (Laurophyllum, Laurus, Rhamnus, Daphnogene, and others). The existence at that time of a dry season (summer drought)
is confirmed by the presence at Terrassa of a 15% of Mediterranean or pre-Mediterranean taxa (*Quercus* cf. *ilex*, *Q. praecursor*). From a physiognomic point of view, 56.4% of the species display leaves with entire margins, while the remainder present dentate or serrated margins. Regarding leaf size, most of the taxa are microphyllous (83.3%) or nanophyllous (10%).

According to these data, the Terrassa association is comparable to floras that are found today in parts of central-east China, South Japan, eastern North America, and North Africa, where a similar mixture of evergreen broad-leaved, warm-temperature, and deciduous elements is present (Wang 1961; Barbero et al. 1982; Richardson 1990; Barbour and Chistensen 1993). In these regions, similar mega-mesotherm taxa (*Lauraceae*, *Myrsine*, *Sapotacites*, *Sapindus*, and others) are concentrated in the lower levels of vegetation, with mean annual temperatures between 16°C and 19°C and mean annual precipitation levels above 1,000 mm. A clear winter season is already present at this stage. In some areas, drier, sclerophyllous elements (*Q. ilex*, *Rhamnus*, *Rhus*) coexist with the former ones. An evergreen broad-leaved forest is present at a medium stage because of the concentration of humidity due to the marine influence (*Cinnamomum*, *Persea*, *Laurus*, etc.). At higher altitudes a deciduous broad-leaved forest (including most of the temperate, deciduous elements like *Acer*, *Fraxinus*, *Juglans*, *Populus*, *Quercus*, *Tilia*, *Zelkova*, and others) is dominant. Mean annual temperatures at this stage decreased to around 12°C. Therefore, a similar zonation probably developed in the transition from the Early to the Late Vallesian, causing a significant decrease of Middle Miocene evergreen elements (33%) and the expansion of a deciduous broad-leaved forest at a medium stage (45% of deciduous elements), where intake of fruit during the winter season must have been much more difficult.

The flora from Terrassa suggests that the deep faunal change at 9.6 Ma was not the consequence of the replacement of the subtropical Miocene forest by grasslands but rather the substitution of one kind of woodland by another. The intensification of the thermal gradients between the middle and low latitudes, probably enhanced by the Himalayan and Tibetan uplifts, led to an abrupt change in the previously existing evergreen subtropical woodlands of western Europe. These were probably replaced by an association in which more seasonally adapted, deciduous trees were dominant. Rather than the moderate cooling associated with the Mi7 isotopic shift, it was this change in the structure of the vegetation which determined the set of extinctions that took place during the Vallesian Crisis.

Most members of this fauna, including *Dryopithecus* and other European hominoids, were mainly frugivorous, with a diet based on fruits and the soft vegetables common in the evergreen broad-leaved forests of the Early and Middle
Miocene. Therefore, although the decrease in temperature and increasing latitudinal gradient had little direct effect on the Vallesian mammals, the replacement of most of the evergreen trees by deciduous ones, well adapted to the new conditions of seasonality with colder winters and dryer summers, had much more dramatic effects. In this way, a number of elements such as certain pigs, rodents, and primates, had to subsist during several months without fruit, a basic and highly nutritional component of their diet. This dietary factor, and not the extension of grasslands or the shift of temperature, was probably the direct agent that caused the abrupt drop of the rich Early Vallesian faunas.

In turn, the sudden disappearance of most of the medium-sized herbivores that had lived in Europe for millions of years probably led to a critical situation for the old predators of Middle Miocene origin, such as the nimravids and the amphicyonids, which until the Early Vallesian had successfully endured the competition from the machairodont cats and the large ursids. These carnivores were also indirect victims of the vegetation change that took place 9.6 Ma. An interesting element is that, contrary to Dryopithecus, the crouzeline pliopithecid Egarapithecus survived the Vallesian Crisis, a fact that is probably related to a folivorous rather than frugivorous diet.

5.7  Asian survivors

The general absence of long sections of Vallesian age makes difficult the recognition of the Vallesian Crisis in other Old World regions, although there exist significant exceptions. In the well-calibrated succession of the Potwar Plateau in central Asia, a significant decay in the relative abundance of tragulids (from 45% to 10% of ruminant artiodactyls; Barry et al. 1991) is observed between 9.8 and 9.3 Ma, while the bovids became the dominant artiodactyls in the area (from 45% to 80%; Barry et al. 1991). However, the changes operating at 9.6 Ma in the region of the Siwaliks can hardly be compared with the dramatic effects of the Vallesian Crisis in western Europe. In contrast with Europe, the mammal faunal association linked to the existence of warm tropical-subtropical forested zones persisted in Siwaliks until chron 4r, at 8.3 Ma. It still included archaic carnivores and rhinoceroses (Brachypotherium), proboscideans (Deinotherium), dormice, shrews, and hominoids with climbing adaptations (Sivapithecus). As we have seen, the persistence of this kind of fauna in the Late Miocene of the Siwaliks can probably be explained by the settling of monsoon atmospheric dynamics in the latter region, which could have maintained the forested subtropical conditions there until 8.3 Ma.
However, between 8.3 and 7.8 Ma, a set of extinctions similar to those of the earlier Vallesian Crisis took place (Pilbeam et al. 1996). Several cricetid, bovid, and tragulid species disappeared, and the hominoid *Sivapithecus* was replaced by colobine monkeys. These faunal changes coincide in the southwestern central Asian region with a shift in the δ¹³C isotopic composition of the paleosoil and dental carbonates, indicating a climatically forced change from a forest and woodland vegetation in which C₃ plants (trees and bushes) were dominant to grasslands dominated by grasses and other C₄ plants (Quade et al. 1989; Morgan et al. 1994; Cerling et al. 1997). They were also largely coeval to further oceanic cooling (White et al. 1997), development of extensive oceanic erosive surfaces (NH6) and development of ice sheets in the Arctic (Eyles 1996). Therefore, the change from C₃ to C₄ dominant vegetation had dramatic effects to the south of the Himalayas and led to the development of an open-woodland also in southwestern Asia.

However, *Dryopithecus* may have subsisted for more time in some refuges of western Eurasia. One of these refuges seems to have been the southern Caucasus, where the presence of a slender dryopithecine (*Udabnopithecus garedziensis*, actually a small form of *Dryopithecus*) was reported a long time ago in the Early Turolian beds of Udabno (Georgia; Gabunia et al. 2001). According to its mammalian association, the age of the Udabno levels is close to 8.5 Ma. It means that *Dryopithecus* or another genus of thin-enamed hominoids persisted in the Black Sea region when this kind of primates went extinct elsewhere in Europe. The persistence of the evergreen subtropical forests to the south of the Caucasus, linked to the retention of special climatic conditions, probably enabled *Udabnopithecus* to survive in this region.

In Asia, the dryopithecines may have survived for longer in the Late Miocene, as suggested by the presence of *Lufengpithecus* in the Late Miocene of Chiang Muang, in Thailand (Chaimanee et al. 2003; Pickford et al. 2004), and Keiyuan and Lufeng in China (Zhang Xingyong 1987; Harrison et al. 2002). Chiang Muang presents a typical pre- *Hipparion* fauna with gomphotherids (*Tetralo- phodon cf. xiaolongtanensis*), rhinos (*Chirotherium intermedium*), peccaries (*Pecarichoerus sminthos*), large suids (*Hippopotamodon cf. hyotherioides*), tetra- conodontine suids (*Conohyus sindiensis, Parachleuastochoerus sinensis*), and tragulids (*Dorcatherium* sp.) in a wooded context that resembles that of the late Middle Miocene of western Europe and Siwaliks. Similar in age and environment is the locality of Keiyuan, in China, which presents the same kind of hominoid (*Lufengpithecus keiyuanensis*). *Lufengpithecus* is still present in type locality of the genus, Lufeng, a site dated in 8 Ma (equivalent, therefore, to the Early Turolian levels of Europe).
5.8 The Oreopithecus fauna: survivors in an island environment

Even more significant was the case of Oreopithecus, an enigmatic hominoid which lived in the Tuscany area (northern Italy) from 9 to 7 Ma. At the time when Oreopithecus occupied the Tuscany region, Italy had a very different aspect from today. The territories which form the present Italian peninsula were in the Early Turolian an arch of isolated islands which extended from central Europe to northern Africa. One of these islands, close to the European mainland, was formed by Tuscany and the Corso-Sardinian block. A number of European immigrants settled in this area at some time between the Vallesian and the Turolian and persisted there until the end of the Miocene. The Oreopithecus faunas appear in several localities from Tuscany, like Casteani, Montebamboli, Ribolla, Montemassi, and also Fiume Santo in Sardinia. However, the best sequence is recorded in the Bacinello Basin, again in Tuscany, where a succession of fossiliferous levels have been recorded (Rook et al. 1999).

The lowermost ones, called V 0 and V 1, are Early Turolian in age and still include some “common,” nonendemic elements like the cricetid Kowalskia and the murids Huerzelerimys and Parapodemus. However, most of the Oreopithecus faunas of the Baccinello levels V 1 and V 2 are basically composed of endemic elements. These faunas appear as a sort of impoverished, “miniaturized” Vallesian ecosystem. Thus, although already modified by the new insular conditions, most of the large mammalian components of this Late Miocene immigration wave can be referred to common elements of the Late Vallesian or Early Turolian European ecosystems, such as hypsodont bovids (Thyrrenotragus, Maremmia), giraffids (Umbrotherium), Microstonyx-like suinae (Eumaiochoerus), dryopithecids (Oreopithecus), or Indarctos-like ursids. Umbrotherium is a poorly known giraffid, probably related to a sivatherine stock. Thyrrenotragus and Maremmia were small bovids with very hypsodont dentitions. Both forms were once interpreted as African immigrants in the area: Tyrrhenotragus as a neotragine (the tribe that includes dwarf antelopes and gazelles) and Maremmia as a precocious alcelaphine (the tribe that includes the African gnus, hartebeests, and impalas). However, some of the features that relate them to these African groups, like the short metapodials of Thyrrenotragus and the probably ever-growing incisors of Maremmia, could have developed independently as specializations linked to an island environment. A small suid, Eumaiochoerus, is also present in the lignites of Bacinello (V 2) and in Montebamboli. It bore a short snout, elongated spatulate upper incisors, and small-sized, chisel-shaped lower tusks. Despite these dental specializations and its small size, other features closely relate this endemic suid to the large Microstonyx major.
However, other elements in the Baccinello succession suggest that a previous Middle Miocene faunal background already existed on the Tusco-Sardinian Island before the Late Miocene settlement of *Oreopithecus* and its allies. This is the case, for instance, of *Anthracoglis*, a dormouse close to the Middle Miocene *Microdyromys* and *Bransatoglis* but significantly larger. A second unnamed giant dormouse is scarcely present in the level V 1 of Baccinello. Besides these endemic dormice, a third small mammal, the lagomorph *Paludotona*, is present in the V 1 level. *Paludotona* was an ochotonid whose body dimensions were again larger than those of its coeval relatives in Europe. The most striking feature of *Paludotona* is its archaic dental morphology, which relates it to some Early to Middle Miocene ochotonids like *Lagopsis*. But the last *Lagopsis* disappeared from Europe in the Middle Miocene, some million years before the deposition of the Bacinello lignites! It seems therefore that at the time of deposition of the V 1 level, there was a long history of isolation on the Tusco-Sardinian Island. The existence of a previous Early to Middle Miocene settlement of the Tusco-Sardinian Archipelago is also supported by the presence in Casteani (equivalent to level V 1 of Bacinello) of an anthracothere. The Tuscan anthracothere is a very archaic one which clearly differs from the advanced Late Miocene anthracotherids of northern Africa. Since the last anthracotheres disappeared from Europe in the Early Miocene, its presence in this area can only be explained as a result of an immigration event from Africa or by assuming its persistence as an Early Miocene relict as in the case of *Anthracoglis* and *Paludotona*.

The environment in which the *Oreopithecus* fauna developed was a mixed mesophytic forest, similar to those that today are found in East Central China along the Yangtze River. Thus, tropical-subtropical trees, like *Engelhardtia*, and warm-temperate trees and shrubs, like *Taxodium*, *Myrica*, etc., are well represented in the pollen analysis of the levels V 0 and V 1 of Bacinello, while more temperate and Mediterranean elements, such as *Quercus*, *Carpinus*, *Tilia*, *Carya*, *Petrocarya*, etc., are rare (Harrison and Harrison 1989; Benvenuti et al. 1994). The pollen analysis developed in the level V 2 shows an increase in temperate, cold-temperate, and mountain elements such as *Picea* and *Abies* (Benvenuti et al. 1994). However, a cold phase cannot clearly be recognized due to the scarcity of grains of elements such as *Tsuga* or *Cedrus*. The increase of temperate and cold-temperate trees in the level V 2 of Bacinello may be the result of a change in the climatic conditions (as in the case of the Late Vallesian floras) but could also be associated with the uplift of the Tuscan area.

The Tusco-Sardinian experiment came finally to an abrupt end when a connection to the continent was established at about 6.5 Ma. New herds of European immigrants entered the Tusco-Sardinian area, including such large predators as *Machairodus* and *Metailurus*. Not surprisingly, *Oreopithecus* and the other
endemic elements of the Tuscan fauna underwent a rapid extinction, unable to resist the competition of the continental newcomers. This change in the faunal composition is also paralleled by a vegetation change. Therefore, the pollen analysis developed in the V 3 unit of Bacinello shows an increase in herbs (Chenopodiaceae, Compistae, Dipsacaceae, etc.) and sclerophyllous trees (Pinus t. haploxylon, Cathaya; Benvenuti et al. 1994). The Oreopithecus experiment came to an end, and the Tusco-Sardinian biome finally followed the general trend toward more open and dry environments that was dominant in the whole of Eurasia.

5.9 The Late Miocene African record

As in other parts of the Old World, the beginning of the Late Miocene in Africa is characterized by the dispersal of the first hipparionine horses on this continent. As happened in Europe, the entry of these hipparionine horses at the beginning of the Sugutan (an African equivalent of the Eurasian Vallesian) did not involve a significant restructuring of the existing terrestrial ecosystems. Most of the elements which composed these ecosystems at the beginning of the Late Miocene in Africa were close relatives of similar taxa in western Eurasia. Faunas of this age, such as those of the Ngororo Formation (Tugen Hills, Kenya; Hill et al. 1985), Narumungule Formation (Samburu Hills, Kenya; Nakaya 1994), or Chorora (Ethiopia; Geraads et al. 2002), are based on a mixture of Middle Miocene African survivors associated with several elements that are common to the Siwaliks and Greek-Iranian provinces. Large browsers include proboscideans (Choerolophodon, Tetralophodon, Deinotherium cf. bozasi), rhinos (Chilotheridium, Paradiceros, Brachypotherium), chalicotheres (Ancylotherium), hipparionine horses (Hipparotherium primigenium), and a variety of ruminants of western Eurasian affinities (Palaeotragus, Samotherium, Protagocerus, Miotragocerus, Palaeoearia/Sivoreas, Homiodorcas, Ouzoceros, Pseudotragus, Pachytragus, Gazella). The persistence of forest conditions is indicated by the presence of tragulids (Dorcatherium pigotti) and gliding rodents (Paranomalurus). However, the presence of an irantherine rhino (Kenyatherium) at Narumungule suggests the existence of more open conditions close to the woodlands. Typical African components of these faunas are the climacoceratid giraffoids (Climacoceras), listriodontine, and tetracodontine suids (Lopholithriodon and Nyanzachoerus), and archaic hippos (Kenyapotamus coryndoni). As happened at the beginning of the Vallesian in Europe, the dispersal of the hipparionine horses involved other Asian elements, such as the saber-toothed machairodontine cats and the “false-hyenas” of the genus Percrocuta. These elements joined other persisting Middle Miocene carnivores of
Eurasian origin, such as the hypercarnivore amphicyonid Agotherium and the ratel-like mustelid Eomellivora.

The persistence of forest conditions in this part of Africa is probably explained on the same basis as in southwestern Asia, that is, as a consequence of the monsoonal dynamics in this region. However, as happened in Siwaliks, a dramatic change is observed between 8 and 7 Ma in the mammalian communities of East Africa. At this time, there is a significant faunal turnover, involving the replacement of close to 75% of the mammal species. Most of the old Middle Miocene holdovers are replaced by close relatives of the elements found today on the modern savannas. Therefore, this time records the first occurrence in Africa of leporids, hominids, new and extant viverrids, extant hienids, new felids, extant and diverse elephantids, new hippopotamids, extant giraffids, and several extant bovids. The former small- to medium-sized browser-based faunas are replaced by a new assemblage in which medium- to large-sized grazers, large browsers, and pursuit carnivores are dominant (Harris 1993).

This change parallels in many ways the one observed at the same time in the Siwaliks, when the former woodland biome opened and C4 grasslands expanded over large parts of western Eurasia. Similarly, the period between 8 and 7 Ma records the expansion of the true savannas in the African continent, leading to the kind of open woodland and grasslands that are present in most of eastern and southern Africa. This is best exemplified by such latest Miocene faunas (Kerian) as Lothagam and Lukeino in Kenya, or those of the Adu-Asa and the Sagantole formations in Ethiopia. At Lothagam, recovery of more than 2,000 identifiable remains has produced a faunal list of more than 30 mammalian species (Leakey and Harris 2003). A number of new browsers replaced the medium-sized community of Middle Miocene origin, largely based on climacoceratid giraffoids and tragulids. Among them we find modern suids of the genera Potamochoerus and Phacochoerus, the peccary-like Cainochoerus, the tragelaphine antelope Tragelaphus, the black rhino Diceros, and the gomphotheres of the genus Anancus. These medium to large browsers joined a number of persisting elements that root in the former Sugutan faunas, such as the suids Kubanochoerus and Nyanzachoerus, the giraffid Palaeotragus, the bovids Tragoportax and Sivatherium, and the rhino Brachypotherium.

However, the most distinctive feature of the new fauna was the sharp increase in medium to large grazers, such as the large sivatherine giraffids of the genus Sivatherium, the first giraffines of the genus Giraffa, the elephantids Stegotetralabelodon, Primelephas, and Elephas, the modern hippos of the genus Hexaprotodon, and the white rhino Ceratotherium. A new form of slender, grazer hipparionine horse, Eurygnathohippus, replaced the former mix-feeder Hippotherium primigenium. But the change to a grazer-dominated community is best exemplified by
the bovids. Therefore, the new faunas include the first and abundant record of antelopes of the tribes Alcelaphini (*Damalacra*), Hippotragini (*Hippotragus*), Reduncini (*Kobus, Menelikia*), Aepycerotini (*Aepyceros*), and Antilopini (*Gazella*), which inhabit the present-day savannas.

The increase in diversity of large browsers and grazers led also to a change in the predator community. While the bear-dog amphicyonids persisted, the number of pursuit carnivores suddenly increased. Thus, at Lothagam are represented several species of cursorial, wolflike hyenas (*Ictitherium, Hyaenictitherium, Hyaenictis, Ikelohyaena*), two species of macharodontine cats (*Lokotunjailurus, Dinofelis*), as well as hunting viverrids (*Genetta*), and mustelids (*Ekoromellivora, Ekorus*).

Therefore, the change operated between 8 and 7 Ma from predominantly closed to predominantly open savanna environments led to a significant change in the mammalian communities, characterized by a dramatic increase in the number of medium to very large grazers, an increase in guild depth of large and very large browsers, and the presence of abundant and diverse pursuit carnivores (Harris 1993).

As evidenced by the Lothagam association, this is the environment where the first hominids succeeded. Besides a number of baboons (*Parapapio, Theropithecus*) and colobine monkeys (*Cercopithecoides*), Lothagam records the presence of an unidentified hominid, represented by some isolated teeth and a fragmented mandible. The site of Lukeino, dated between 6.2 and 5.6 Ma, also delivered remains of a hominid, *Orrorin tugenensis*, in a context that resembles that of Lothagam, with elephants (*Primelephas, Stegotetrabelodon, Loxodonta*), gomphotherids (*Anancus*), deinotheres (*Deinotherium*), rhinos (*Ceratotherium, Diceros*), chalicotheres (*Ancylotherium*), hippparions (*Eurygnathohippus*), hippos (*Hippopotamus, Hexaprotodon*), suids (*Nyanzachoerus*), giraffids (*Giraffa*), and a variety of antelopes: Cephalophini (*Cephalophus*), Reduncini (*Kobus*), Aepycerotini (*Aepyceros*), Tragelaphini (*Tragelaphus*), and Neotragini (Pickford and Senut 2001). The first hominids of the genus *Ardipithecus*, from the latest Miocene (5.7–5.5 Ma) of the Adu-Asa and Sagantole Formations in the Middle Awash, lived in a somewhat different environment (Wolde Gabriel et al. 2001). Although, as in Lothagam and Lukeino, large grazers are present (the elephants *Primelephas* and *Stegotetrabelodon*, the giraffid *Sivatherium*, the reduncine antelope *Kobus*, as well as hipparions), the dominance of browsing forms such as gomphotherids (*Anancus*), deinotheres (*Deinotherium*), rhinos (*Diceros*), hipppos (*Hexaprotodon*), suids (*Nyanzachoerus*), and boselaphine and tragelaphine bovids (*Miotragocerus, Tragelaphus*) points to more wooded conditions.

Different again was the association found at Toros-Menalla 266, in Chad, dated between 7.4 and 5.2 Ma, where another early hominid species,
Sahelanthropus tchadiensis, was described (Vignaud et al. 2002). The faunal assemblage of Toros-Menalla shares with Lothagam and Lukeino the presence of a variety of large grazers: antelopes (Kobus, Hippotragini, Antilopini), elephants (Loxodonta), giraffids (Sivatherium), and hipparions (Eurygnathohippus cf. *abudhabiense*). However, the remaining association is based on medium to large browsers or mix feeders: Anancus, Nyanzachoerus, Hexaprotodon, aff. *Palaeoryx*. Particularly surprising is the presence of the large, advanced anthracothere *Libycosaurus petrochii* (also referred as *Merycopotamus petrochii*), an element which is absent from the faunal associations of the same age in the eastern African Basins. In fact, the Toros-Menalla 266 faunal assemblage fits better with the association found at the site of Sahabi (Libya), south of Bengazhi.

The fauna from the Sahabi Formation (Boaz et al. 1987) shares a number of elements in common with the eastern basins, such as monkeys (*Macaca, Libypithecus*), archaic elephants (*Stegotetrabelodon*), rhinos (*Diceros*), hipparions (*Eurygnathohippus*), hippos (*Hexaprotodon*), tetraconodontine suids (*Nyanzachoerus*), giraffids (*Samotherium*), and a variety of antelopes: Reduncini (*Redunca*), Alcelaphini (*Damalaca*), Hippotragini (*Hippotragus*), and Antilopini (*Gazella*). However, a significant part of the association is composed of taxa that were common to the Greek-Iranian province and the Siwaliks region. This is the case, for instance, with the above mentioned anthracotherid *Libycosaurus* (related to *Merycopotamus*), the bovids *Miotragocerus*, *Prostrepsiceros*, and *Leptobos*, the amebelodontid gomphothere *Amebelodon*, and the whole carnivore taxocoenosis (*Macairoidus, Chasmaportetes, Hyaenictitherium, Adrocuta, Indarctos, Agriotherium*; Howell 1987). In Sahabi, the botanical evidence provided by well-preserved fossil wood indicates an environment dominated by wooded savanna and semidesert grassland, with dominance of *Acacia* (64%) and Mimosaceae (Dechamps and Maes 1987). The existence of periodic fires (a usual phenomenon in a savanna environment) is recorded as traumatic rings in fossil wood. The estuarine context of this site is reflected by the relative high proportion (13.5%) of Salicaceae (*Populus euphratica*).

The western-Asian character of the Sahabi fauna, which partly extends south to the Chad Basin (Toros-Menalla 266), opens the question of the origin of the Late Miocene African faunas, after the crisis at 7 Ma. It has been suggested that much of the current savanna fauna did not evolve in situ from the Early and Middle Miocene African mammals but migrated from more northern latitudes in the Late Miocene, replacing the previous forest endemic dwellers (Maglio and Cooke 1978). More specifically, with the drying out of Africa between 8 and 7 Ma, large mammals from the Greek-Iranian province may have colonized the lower latitudes, their adaptations to a sclerophyllous woodland (hypsodont teeth, cursorial skeletons) having acted as exaptations (*sensu* Gould and Vrba 1982)
to a savanna biome (Solounias et al. 1999). This was probably the case for bovids, giraffids, equids, hyaenas, rhinos, and hominids. The faunal composition of sites like Toros-Menalla and Sahabi strongly supports this scenario.

References

Continental-scale hypsodonty patterns, climatic paleobiogeography and dispersal of Eurasian Neogene large mammal herbivores. Deinsea 10: 1–11


Harrison TS, Harrison T (1989) Palynology of the late Miocene Oreopithecus-bearing lignite from Baccinello, Italy. Paleogeogr Paleoclim Paleoloc 76: 45–65


Nakaya H (1994) Faunal change of Late Miocene Africa and Eurasia: mammalian fauna from the Namurngule Formation, Samburu
Hills, Northern Kenya. Afr Study Monogr 20(Suppl.): 1–112


Richardson SD (1990) Forest and forestry in China. Island Press, USA


6 Postcranial and Locomotor Adaptations of Hominoids

Carol V. Ward

Abstract

Extant apes are adapted to various forms of below-branch forelimb-dominated arboreal locomotion and share morphologies associated with the shared aspects of their locomotor behaviors. With the expanding record of Miocene hominoid fossils, we are coming to realize that although some shared characters may indeed be homologous, at least some almost certainly represent homoplasies. The apparently more primitive body plan of Sivapithecus than seen in Asian and African great apes indicates that at least some homoplasy has occurred within these clades. Furthermore, the expanding fossil record may be indicating a greater diversity of positional behaviors within the Hominoidea than previously appreciated; for example, Pierolapithecus may indicate the evolution of suspensory locomotion in combination with arboreal quadrupedalism, and Nacholapithecus is unique with its enlarged forelimb but otherwise primitive body plan. These new fossils reveal that variation is prevalent and critical to appreciate for reconstructing hominoid evolutionary history. Furthermore, it seems increasingly likely that many postcranial and locomotor specializations of great apes may have evolved from ancestors that were more generalized than are living hominoids. This realization is critical for interpreting the ancestral morphology from which hominins were derived.

6.1 Introduction

One of the most distinctive shared characteristics of modern apes is their locomotor specialization for below-branch forelimb is dominated arboreal locomotion, many adaptations for which are still seen in humans today (Huxley 1863). Extant apes all exhibit an overlapping set of adaptations to suspension and orthograde climbing in their torsos, limbs, hands, and feet. They have elongated forearms and hands, high-intermembral indices, laterally oriented shoulder joints, limb joints adapted to loading in a variety of postures, short, broad torsos, and absence of a tail. Given this suite of apomorphies, parsimony would seem to
dictate that ancestral apes should have been suspensory as well (Gregory 1916). If not, extant hominoids must have developed these adaptations independently, which would, on the face of it, seem to involve an unlikely amount of homoplasy.

However, while below-branch forelimb-dominated arboreality distinguishes all modern ape taxa from other catarrhines (Schultz 1930, 1969a and b), it is not a monolithic adaptation (Larson 1988). Young (2003) has suggested that great ape postcranial adaptations are homologous (see also Pilbeam 2002) and that hylobatids may be autapomorphic, due to their small size and exclusively arboreal specialization for brachiation that involves moving rapidly through the trees in ricochetal fashion. Yet arboreal adaptations actually vary among extant ape genera more than is often emphasized in considerations of the evolution of hominoid locomotion, reflecting specialization on different types and amounts of climbing and suspensory behavior. Orangutans almost never use the ground, yet, given their large body size, they move cautiously through the upper levels of the canopy with quadrumanous climbing and arm-hanging. Their forelimbs are astonishingly long and all joints are flexible and employed in diverse postures (MacKinnon 1974; Cant 1987).

Chimpanzees, arboreal when foraging, hunting, and sleeping, spend much of their time as terrestrial knuckle walkers (Goodall 1965; Hunt 1990), as do bonobos, although the locomotor behavior of the latter is less well studied (Doran 1993). Gorillas are extremely large and are even more terrestrial than chimpanzees or bonobos, yet particularly in certain habitats they habitually climb trees frequently for feeding and sleeping (Tuttle and Watts 1985; Remis 1995). Hominins, of course, modified their locomotor skeletons to specialize in the most distinctive of all primate locomotor behaviors, habitual terrestrial bipedality, although they likely evolved from ancestors that were at least somewhat adapted to climbing or suspensory arboreal locomotion. Given this variation among taxa, a certain level of independence in the acquisition of some suspensory traits should perhaps not be considered entirely unlikely.

The burgeoning fossil record is strengthening the hypothesis that climbing and suspensory adaptations developed in mosaic fashion over evolutionary time, and occurred in different ways and even multiple times in separate hominoid lineages (Sarmiento 1987; Pilbeam et al. 1990; Moyà Solà and Köhler 1995; Begun et al. 1997; Finarelli and Clyde 2004; Moyà Solà et al. 2004). The hominoid fossil record now includes over 25 genera known from as early as 20 Ma, and many taxa are known from postcranial elements (chapters in Hartwig 2002). Among these fossil taxa, there is much wider range of known locomotor modes than among extant ones. Adaptations seen in some, and even all, extant apes do not occur as a block. For example, Nacholapithecus has elongated forelimbs, but a long, narrow torso and apparently pronograde quadrupedal posture (Rose et al. 1996; Senut
Pierolapithecus has what appears to be a hylobatid-like torso structure, yet does not seem to have particularly long digits (Moya` Solà et al. 2004). Considering variation within living and fossil hominoids may lead us away from dichotomous views on whether the euhominoid ancestor was “great ape-like” or “basal hominoid-like” (Pilbeam et al. 1990; Young 2003), and may lead us to a more nuanced and more accurate understanding of how hominoids evolved.

The positional behavior variation among extant and fossil hominoids provides an important set of information about hominoid phylogeny and evolution. This chapter will summarize the postcranial adaptations of extant and the best-known fossil hominoid genera, and put these taxa into phylogenetic context in order to explore evolutionary patterns in hominoid postcranial and locomotor adaptations.

### 6.2 Extant hominoids

Living apes all share a suite of adaptations to below-branch, forelimb-dominated arboreality, although they do not exploit this locomotor niche identically and are distinguishable postcranially, although particularly among great apes differences are poorly characterized (Young 2003; but see Larson 1988; Inouye and Shea 1997; Drapeau 2001). All, though, have evolved to negotiate terminal branches of arboreal substrates by hanging below branches and distributing their weight among multiple supports. Their relatively large body sizes and long forelimbs allow them to bridge gaps in the canopy rather than leaping. They lack tails, which are no longer necessary for balance within this type of locomotor regime. As a consequence, the musculature usually associated with tails in nonhominoid primates has become restructured to form a muscular pelvic floor, providing support for the viscera during the orthograde postures common in all ape species today (Elftmann 1929).

Apes have high intermembral indices (Schultz 1930; Aiello 1981; Jungers 1985). When climbing, their long forelimbs and relatively short hindlimbs improve their ability to negotiate large in diameter vertical supports. The long arms also assist in increasing their reach for bridging gaps and grasping branches. Limb proportions, vary however (Aiello 1981; Jungers 1984, 1985). Orangutans and hylobatids have proportionally longer forearms and hands than do chimpanzees, and chimpanzees and bonobos alone have differentially elongated metacarpals more than other extant apes (Drapeau 2001).

Ape torsos are broad, with the scapulae positioned dorsally and glenoid fossae oriented laterally (Keith 1923; Schultz 1930, 1961, 1969a; Benton 1965, 2003; Ishida et al. 2004).
Their ribs have higher costal angles, sterna are broad, scapulae have a cranially oriented glenoid fossa, and their clavicles are long (Schultz 1937; Cartmill and Milton 1977; Larson 1993). The broad torso and wide pectoral girdle results in a shoulder position that places the scapulothoracic musculature more in the coronal plane, facilitating adduction of the upper limb (Erickson 1963; Benton 1965, 1976; Ward 1993). Adduction is accomplished primarily by the latissimus dorsi muscle along with others that are important in hoisting the body up from an arm-hanging position, and in pulling the body among supports in the trees (Swartz et al. 1989; Hunt 1991). Extant hominoid pectoral girdle form also increases reach of the forelimbs and circumduction at the shoulder (Cartmill and Milton 1977).

Extant hominoid lumbar spines are reduced in length, with only 5–6 lumbar vertebrae in lesser apes, compared with the typical primate number of 7, and great apes have only 3–4, so that the lower ribs approximate the iliac crest (Schultz and Straus 1945; Erikson 1963; Ankel 1967, 1972; review in Ward 1993). African apes differ from humans, orangutans, and hylobatids in having 13 rather than 12 thoracic vertebrae and ribs (Schultz 1930, 1969a, b). This variation in thoracic vertebral count among extant hominoids is noteworthy and may be yet another small line of evidence supporting hypotheses of postcranial homoplasy among extant apes. All great ape lumbar spines are almost completely inflexible (Slijper 1946; Schultz 1969a, b; Benton 1976). Their iliac blades are craniocaudally elongated, expanding the distance between hip and sacroiliac joints (Waterman 1929; Ward 1993). Part of latissimus dorsi inserts directly upon the iliac crest (Sonntag 1923, 1924; Waterman 1929; Gregory 1950). The reduced lumbar spine provides a stiffer platform for these muscles to act upon the upper limb and protects the vulnerable lumbar spine from excessive lateral bending moments during latissimus dorsi contractions (Cartmill and Milton 1977; Jungers 1984; Sarmiento 1985; Ward 1993).

Reaching overhead during climbing requires the ability to adduct the wrist, which apes can do due to a reduced ulnar styloid that has lost contact with the carpus (Lewis 1971, 1972, 1989; Sarmiento 1988; Whitehead 1993). The pisiform is more distally situated, facilitating extensive adduction at the wrist up to 90° (Sarmiento 1988). Apes rely on a hook grip during suspensory behaviors and have relatively long fingers with strong flexor musculature and reduced thumbs (Schultz 1930; Rose 1988). They also have strong grasping pollices and halluces, along with well-developed musculature for adduction. The fibula, site of attachment of the long hallucal flexor muscles, is well developed (Schultz 1930; Sonntag 1923, 1924; Gregory 1950; Swindler and Wood 1973). Manual and pedal phalanges have well-developed attachment sites for digital flexors (review in Begun 1994a). The curved forearm bones of apes have been linked to the presence of strong hand and finger flexors (Miller 1932; Knussmann 1967).
These adaptations are better developed in great apes than in lesser apes (review in Young 2003). Mechanical constraints on mechanical function, and stresses on the musculoskeletal system, increase exponentially with increasing body size; thus some differences between lesser and great apes may represent allometric issues (Aiello 1981; Jungers 1984). The small body size of hyllobatids may be secondarily derived, as is their unique adaptation for ricochetal brachiation (Cartmill 1985). Their forelimbs and hands are particularly elongate (Jungers 1984), their thumbs particularly short, and their body mass relatively low. Their lumbar vertebral columns are 5–6 segments long and their thoracic columns 12 (Schultz and Straus 1945).

Differences between African apes and orangutans have also been noted, including a smaller supraspinous region of the scapula, smaller acromion processes, and shorter scapulae in orangutans (Oxnard 1984; Larson 1988; Young 2003). Orangutans have higher intermembral indices (Aiello 1981), a nonuniformly curved iliac crest, and an anterior longitudinal ligament ridge on the anterior surfaces of the vertebral bodies. They also lack a ligamentum teres of the femoral head, along with particularly mobile knee and midtarsal joints. Orangutans alone among large hominoids retain separate centrale and scaphoid bones rather than the fused os centrale seen in extant African apes. These differences are presumably related to the highly developed and nearly exclusively quadrumanous arboreal locomotor behavior of orangutans on one hand compared with the partly terrestrial knuckle-walking habits of African apes on the other. It is worth noting, however, that little work has been done to explore the extent of morphological variation among great apes and their functional significance.

So, although apes do share a distinctive suite of morphologies, they are neither identical in their behavior nor anatomy. Therefore, although a subset of shared morphologies undoubtedly was present in their common ancestor and represents synapomorphies of the hominoid clade, such differences indicate a combination either of locomotor specialization within each lineage and/or independent acquisition within lineages undergoing broadly similar selective pressures. Only careful examination of the currently known fossil record, and discovery of new hominoid fossils, will allow us to determine which scenarios led to the diversity of known hominoid locomotor adaptations.

6.3 Fossil hominoids

Hominoids can essentially be grouped into temporal and adaptive grades, basal hominoids and euhominoids (Begun et al. 1997) (Figure 6.1). Basal hominoids are stem taxa known from the Early and Middle Miocene, and comprise taxa that are not crown hominoids and that lack postcranial specializations for extant
apelike below-branch arboreality. Euhominoids are members of a crown hominoid clade, and so demonstrate extant hominoid synapomorphies, of which the postcranial anatomy is significant to this discussion. In our analyses of relatively well-known taxa (Begun et al. 1997; Begun and Ward in press), basal hominoid taxa include Proconsul and Afropithecus from the Early Miocene, and Kenyapithecus, which in our analysis includes Equatorius (Ward et al. 1999) and Griphopithecus (Begun 2000) from the Middle Miocene, and to which we would now add the more recently discovered Nacholapithecus from the Middle Miocene (Ishida et al. 1984, 2004; Rose et al. 1996; Nakatsukasa et al. 1998). Euhominoids include Sivapithecus, Oreopithecus, Pierolapithecus, and Dryopithecus.

### 6.3.1 Proconsul and Afropithecus

Postcranially the best-known genus of Early Miocene hominoid is Proconsul, although Afropithecus is also represented by several postcranial elements. While these taxa differed craniodentally, Afropithecus is strikingly similar in preserved postcranial bones to Proconsul nyandae (Ward 1998). These taxa, among others less well known postcranially, represent a stem group referred to as basal
hominoids (Begun et al. 1997). Comprehensive reviews of these taxa can be found in Walker (1997) and Leakey and Walker (1997).

*Proconsul* was a genus composed of at least four species of pronograde, quadrapedal. Arboreal frugivores that ranged in size from that of colobus monkeys to that female gorillas. *Proconsul* individuals were very generalized animals. They were certainly an above-branch arboreal quadrupeds that would have been capable, as most primates are, of limited below-branch postures and movement, but that show no specialization for below-branch arboreality. *Proconsul* had a roughly even intermembral index and retained ulnar contact with the wrist (Beard et al. 1986). It also had a relatively long torso with six lumbar vertebrae, lacking the stiffening of the lumbar spine seen in extant apes (Ward 1993; Ward et al. 1993). The vertebrae have transverse processes arising from the vertebral body as in most monkeys, distinct accessory processes that suggest large erector spinae musculature, reflecting a narrow rib, cage, and narrow and laterally facing iliac blades (Ward 1993). The pelvis was narrow, as, presumably, was the thoracic cage. The humerus is retroflexed with little torsion, reflecting ventrally oriented glenohumeral joints (Napier and Davis 1959; Larson 1988). The sternebrae are wider than those of cercopithecids, instead resembling atelines, but were not as broad as those of apes.

*Proconsul* is distinguished postcranially from earlier generalized catarrhines such as *Aegyptopithecus* (Rose 1997; Walker 1997) by a shoulder adapted to a wider range of loading, an elbow joint emphasizing joint loading in a variety of flexion–extension and pronation–supination postures, longer distal limb segments, a foot that may have been better adapted to inversion-eversion (Walker 1997: table 1 p 221), and by the lack of a tail (Ward et al. 1991; Nakatsukasa et al. 2004). It is also true that *Proconsul* had a hip joint with a high neck-shaft angle, a femoral head set high on the neck, a centrally placed fovea capitis, and a greater trochanter that was shorter than that of many monkeys indicating a hip joint adapted to loading in at least somewhat abducted postures (Ward 1992). The knee joint was broad, with a broad, flat patella, also indicating a very generalized and adaptable use of the lower limb. Most monkeys and earlier, generalized catarrhines have joints designed for fairly stereotypical loading environments principally comprising of quadrupedal postures in which the limbs are relatively adducted, the hands are pronated and palmigrade. Thus in comparison to *Aegyptopithecus*, *Proconsul* and other known Early Miocene basal hominoids were adapted for using their limbs in more abducted postures, with hands and feet grasping arboreal supports in a variety of positions.

The hands and feet of *Proconsul* have relatively well-developed first rays and supporting bones such as the large fibula. The phalanges are fairly long, with clear ridges for attachment of the finger and toe flexors (Begun 1993). The hallux and
pollex were long relative to the other digits, suggesting that actual grasping was used rather than emphasizing the hook grips characteristic of extant apes.

Before the discovery of taillessness in *Proconsul*, reduction of a tail and commensurate restructuring of the pelvic floor musculature had been equated with the evolution of orthograde posture, which requires muscular support of pelvic viscera. However, the lack of a tail in *Proconsul*, and also in *Nacholapithecus* (Ward et al. 1991; Nakatsukasa et al. 2004), reveals that tail loss in hominoids was not associated with orthogrady, but rather with a decreased reliance on running and leaping behaviors and an emphasis on more deliberate arboreality, emphasizing manual and pedal grasping with weight distributed over multiple supports (Kelley 1995). Taillessness was likely a characteristic of all basal hominoids and one of the defining features of this superfamily.

In summary, a reasonable locomotor reconstruction of early hominoids is possible. The positional repertoire of the earliest known hominoids likely was distinguished from that of primitive catarrhines such as *Aegyptopithecus* by a specialization for deliberate arboreal climbing and clambering, using strong manual and pedal grasping to maintain balance and to move about in the arboreal environment. The emphasis on grasping meant that tails were no longer needed for balance, and thus an external tail became lost. A grasping adaptation enabled many hominoids to attain larger body sizes than seen in extant monkeys, yet still remain arboreal. This emphasis on cautious climbing and clambering enabled basal hominoids to reach fruit on the terminal branches and cross gaps in the canopy by giving them the ability to distribute their weight over multiple supports. They almost certainly also practiced vertical climbing and limited suspension but had not yet evolved selection for specialized adaptations to these behaviors; instead they spent most of their time above the branches. Their limb joints were adapted to loading in a wider variety of postures than is typical for most nonhominoids, including those involving an abducted hip and supinated elbow. They were certainly capable of orthograde postures and probably employed them often. Yet their narrow torso structure with ventrally oriented shoulder joints and ulnocarpal contact that limited ulnar deviation of the wrist indicate locomotion primarily with the limbs positioned mainly underneath the body, or slightly abducted.

This generalized arboreal locomotor habitus permitted this early radiation of basal hominoids to exploit a variety of body sizes, as well as a breadth of ecological niches. *Proconsul* was likely a ripe-fruit frugivore and *Afropithecus* a seed predator (Leakey and Walker 1997). *Rangwapithecus*, less known postcranially, seemed to be at least partly folivorous (Kay and Ungar 1997). Thus, except for *Morotopithecus* (below), Early Miocene apes seem to be an adaptive radiation that shared a similar set of locomotor adaptations.
6.3.2 *Morotopithecus*

*Morotopithecus bishopi* is the earliest Miocene ape known from eastern Uganda and dated at over 20.6 Ma (Gebo et al. 1997). The palate of this taxon resembles those of other basal hominoids, and has been compared to both *Proconsul* (*P. major*) and *Afropithecus* (review in Leakey and Walker 1997). As such, this species is placed phylogenetically as a basal hominoid rather than as an euhominoid. Postcranially, however, it differs from other Early Miocene apes in displaying a more derived, modern apelike postcranial skeleton in some respects. The only postcranial elements known from *Morotopithecus* include a glenoid fossa, a lumbar vertebra and associated fragments of other lower vertebrae (Walker and Rose 1968), and partial femora.

The *Morotopithecus* glenoid fragment is similar to those of extant great apes, being ovoid and shallow and lacking a narrow, laterally concave cranial portion (MacLatchy et al. 2000). This suggests a shoulder joint adapted to loading in a variety of postures, and may be associated with a broader upper torso morphology. A more extant-apelike torso in *Morotopithecus* compared with *Proconsul* is also suggested by the lumbar vertebrae (Ward 1993b; Sanders and Bodenbender 1994), which are morphologically similar to those of hylobatids in that the lumbar vertebral transverse processes which arise from the junction of the vertebral body and pedicle (Walker and Rose 1968) rather than from the body as in *Proconsul* (Ward 1993; Ward et al. 1993) or *Nacholapithecus* (Nakatsukasa et al. 1998). This is correlated with the lack of accessory processes (Ward 1993). The body is also not hollowed out laterally, although a median keel is present. The *Morotopithecus* femur is robust with a small head but displays the typical hominoid pattern of having a high neck-shaft angle and centrally placed fovea capitis. The distal femur is very broad mediolaterally, with condyles asymmetric in size, and there is a large popliteus groove as in extant apes, suggesting adaptation to habitual loading in a variety of postures, and a high degree of rotation of the knee joint.

Altogether *Morotopithecus* shares several key features with extant apes, suggesting adaptation for more below-branch arboreal activities. The shoulder, hip, and knee are adapted to a variety of postures, and the vertebrae suggest a certain amount of broadening and stiffening of the torso. However, there is nothing in the *Morotopithecus* skeleton that is substantially more derived than that of hylobatids except for a broader knee joint and the presence of a gluteal ridge, and perhaps also transverse process inclination (Young and MacLatchy 2004). The femoral characters may be related to body size variation, and the transverse process inclination is actually somewhat intermediate between that of gibbons and great apes (Shapiro 1993). Still, the hominoid similarities have led MacLatchy
(2004), MacLatchy et al. (2000), and Young and MacLatchy (2004) to hypothesize that *Morotopithecus* is a basal member of the euhominoid clade and that other basal hominoids, like *Proconsul*, *Afropithecus*, *Nacholapithecus*, and *Equatorius/Kenyapithecus* at least, are more distantly related to extant apes. This hypothesis assumes that the postcranial adaptations shared by extant hominoids in torso restructuring, below-branch locomotion, and knee mobility are synapomorphies.

Given its primitive craniodental characters, the phylogenetic placement and thus the evolutionary implications of *Morotopithecus* are difficult to assess. But, considering the homoplasic similarities between ateline monkeys, *Pliopithecus* and extant hominoids, and the variation among extant hominoids, it certainly seems plausible that shared features of *Morotopithecus* and great apes represent homoplasies. Regardless of its phylogenetic affinities, it seems apparent that *Morotopithecus* was a more forelimb-dominated below-branch arboreal animal, with a greater emphasis on orthograde postures than is typical for other basal hominoids.

### 6.3.3 *Nacholapithecus*

*Nacholapithecus kerioi* is a Middle Miocene ape from northern Kenya (Ishida et al. 1984, 2004; Pickford et al. 1987; Rose et al. 1996; Nakatsukasa et al. 1998). It displays an unusual suite of morphologies in a very different combination than known for other hominoids. Its overall body plan is similar to that of *Proconsul*. It has a long vertebral column with six lumbar vertebrae shaped like those of *Proconsul* (Nakatsukasa et al. 1998). It appears to have had a narrow torso shape and no tail, yet its feet were long, shoulder joint mobile and upper limbs very large (Nakatsukasa et al. 1998; Senut 2003; Ishida et al. 2004). The hands and feet reflect palmigrade, plantigrade quadrupedalism.

The body proportions of *Nacholapithecus* are striking, having very large forelimb and relatively small hindlimb elements, outside the observed ranges for even extant apes (Ishida et al. 2004). The lower limb is comparatively small, but there is a high neck-shaft angle of the femur and a short neck. Otherwise, the tibia, fibula, and tarsal bones are similar to those of *Proconsul*. What is different are the very long pedal digits found in *Nacholapithecus*, which are similar in size to those of chimpanzees despite the fact that it is thought to have weighed about half of what chimpanzees weigh (22 kg; Ishida et al. 2004). The distal humerus is derived relative to that seen in *Proconsul*, with a better-developed ball-and-socket morphology of the humeroradial joint and a more symmetrical humeroulnar joint. It had a relatively long clavicle with ligament markings, suggesting that protraction of the humerus into overhead postures would have been emphasized.
rather than abduction (Senut 2003). The morphological similarity between the shoulder bone morphologies and those of colobine monkeys underscores this interpretation. This difference in inferred upper limb use in climbing between *Nacholapithecus* and euhominoids is likely related to its lack of torso restructuring from the primitive condition. This restructuring to make the torso shorter, stiffer, and broader is seen in apes emphasizing abduction–adduction movements of the upper limb (Ward 1993). Thus, *Nacholapithecus* does seem to have been adapted for more extensive forelimb-dominated climbing locomotion than were earlier basal hominoids, with enlarged forelimbs and small hindlimbs and long pedal phalanges, but in a fundamentally different way from what is seen in later euhominoids.

### 6.3.4 *Equatorius*/Kenyapithecus

The Middle Miocene African fossil sample originally attributed to *Kenyapithecus* has now been partly divided by some researchers, and placed into *Equatorius* (Ward et al. 1999) or *Griphopithecus* (Begun 2000), with the result that there is now only one humerus known for *Kenyapithecus wickeri* (summary by Ward and Duren 2002). Given the taxonomic debate, here all samples are treated together. *Equatorius* resembles other basal hominoids in several respects (Sherwood et al. 2002). It has a monkey-like scapula and a retroflexed humeral shaft, indicating a pronograde posture at the shoulder. The ulna still contacts the carpus, and the os centrale is unfused to the scaphoid. *Equatorius* has the mobile hip and knee joint morphology characteristic of all hominoids, and slightly broader sternebrae than typical for monkeys, although it is reconstructed to lack the laterally expanded torso of extant apes. *Equatorius* (=*Kenyapithecus*) from Maboko Island has been interpreted as semiterrestrial based on its retroflexed humeral shaft, medial humeral epicondyle, and olecranon process, along with long radial head, short phalanges, and morphology of the metacarpal heads (McCrossin and Benefit 1997). However, it has also been described as resembling extant apes in having a straight humeral shaft and strong grasping capabilities (McCrossin 1997). It is difficult to reconcile these divergent descriptions at present. *Equatorius* does not display evidence of thoracic reorganization typical of *Morotopithecus* or extant apes, but instead was a more generalized pronograde quadruped that may have spent some time on the ground. If indeed *Equatorius*/Kenyapithecus is more closely related to euhominoids than is *Proconsul* (Ward and Duren 2002), it reinforces the hypothesis that the ancestral euhominoid was generalized postcranially.
6.3.5 *Sivapithecus*

The discovery of humeral shafts attributed to *Sivapithecus* (Pilbeam et al. 1990) led to a startling revelation about the evolution of hominoid locomotion, referred to as the “*Sivapithecus* dilemma” by Pilbeam and Young (2001). The hand and foot remains known for *Sivapithecus* up to that time all exhibited a more derived modern apelike morphology than earlier hominoids such as *Proconsul*. It had longer, stronger phalanges, a more well-developed hallux, and a more mobile elbow, and these morphologies were presumed to indicate an extant great-ape-like overall postcranial gestalt. To most researchers the marked craniofacial similarities between *Sivapithecus* and *Pongo* indicate a close phylogenetic relationship (Ward 1997), so the apparently modern apelike postcranial fossils fit a phylogenetic position of *Sivapithecus* well within the hominid clade.

The humeri, however, appear to paint a different picture. Although missing their proximal ends, both specimens appear to be inclined posteriorly and medially, the morphology seen in primates like most monkeys, *Proconsul* and *Nacholapithecus* that have posteriorly directed humeral heads, ventrally facing scapular glenoid fossae, and narrow torso structures.

If *Sivapithecus* had a narrow torso, and yet was a member of the pongine clade, this means that many of the postcranial apomorphies shared among *Pongo*, *Pan*, and *Gorilla* must represent homoplasies rather than homologies as previously interpreted. *Pongo* must have evolved a similar torso structure, limb proportions, and other below-branch specializations independently from African apes. This presented paleoanthropologists with the initially uncomfortable realization that the unique locomotor mode essentially shared by extant apes evolved independently at least twice.

6.3.6 *Pierolapithecus*

The hypothesis that postcranial homoplasy has occurred among extant hominoids is bolstered by the recent discovery of *Pierolapithecus catalaunicus* from the Middle Miocene of Spain (Moya´ Sola´ et al. 2004). Although this form was originally interpreted as a basal hominid, it appears most parsimonious to consider *Pierolapithecus* a member of the hominine (African ape and human) clade (Begun and Ward, 2005). The vertebrae of *Pierolapithecus* resemble those of *Morotopithecus*, and among extant apes, are most like those of hylobatids. It had a high costal angle, revealing that its torso shape was also hominoid-like. It lacked ulnar contact with the carpus, indicating the ability for substantial ulnar deviation. *Pierolapithecus* also lacks some shared specializations of extant great apes,
such as dorsally positioned vertebral transverse processes, elongate phalanges, and some reorganization of the carpus. Too few elements are known from both *Sivapithecus* and *Pierolapithecus* to make extensive direct comparisons, so it is hard to say if they represent an ancestral morphology from which all extant apes evolved. But, because it lacks shared specializations of extant great apes, *Pierolapithecus* provides more evidence that *Pongo* is convergent upon African apes in its specialized below-branch arboreal locomotor skeleton.

### 6.3.7 Oreopithecus

*Oreopithecus bambolii* is known from most skeletal elements and has been the subject of intense debate regarding its locomotor behavior (Harrison 1986; Sarmiento 1987; Harrison and Rook 1997; Pilbeam 2004; Moyà Solà et al. 1999; Rook et al. 1999). It has even been considered to be bipedal (Köhler and Moyà Solà 1997). It has an extremely high intermembral and brachial indices, resembling only those of orangutans. It has five lumbar vertebrae, as do humans and hylobatids, and a relatively long, broad pelvis that is broader than those of hylobatids. Its upper thoracic and pectoral girdle morphology is more difficult to assess due to the crushing experienced by the published fossil bones. *Oreopithecus* also has a relatively large femoral head relative to neck size, associated with extensive hip joint mobility. It was clearly adapted for below-branch arboreality, although its metacarpals and phalanges are not particularly long (Moyà Solà et al. 1999; Rook et al. 1999), perhaps suggesting that manual grasping was done between the pollex and medial digits rather than with the hook grip typical of extant apes. This would be yet another example of variations on the below-branch forelimb-dominated theme. *Oreopithecus* also demonstrates that hominoids can specialize in this type of locomotor behavior without having every specialization seen within any given extant ape taxon.

### 6.3.8 Dryopithecus

*Dryopithecus* sp. have the most derived postcranial skeleton of any Miocene ape (Begun 1993, 1994b; Moyà Solà and Köhler 1996; Begun et al. 1997), except perhaps for the putative hominin *Orrorin tugenensis* (Senut et al. 2001). *Dryopithecus* has dorsally positioned lumbar transverse processes that arise from the neural arch and are inclined posteriorly. It is not possible to determine how many were originally present. It has long, powerful manual phalanges coupled with short metacarpals, and limb proportions similar to those of modern chimpanzees.
and bonobos (Moya Solà and Köhler 1996). The large lunate and other carpals suggest reduced ulnar–carpal contact (Begun 1994b; Moya Solà and Köhler 1996). The vertebral morphology indeed reflects a great apelike pelvic structure and is likely related to thoracic and pectoral girdle form as well. In this case, Dryopithecus shares many features with extant great apes to the exclusion of hylobatids, and since it is universally recognized to be a euhominoid, suggests that at least some form of below-branch arboreality characterized common ancestors of living euhominoids.

6.4 Locomotor evolution in hominoids

Putting the fossil evidence together, it appears that an emphasis on deliberate arboreal quadrupedalism with weight support over multiple branches, hereafter referred to as clambering, probably served as a preadaptation for the below-branch specializations of later hominoids, as well as for Morotopithecus. Once the tail was lost, the tail musculature was in a position to become a strong pelvic floor, facilitating more extensive orthogrady. The strong manual and pedal grasping of multiple supports, and relatively mobile limb joints, provided a platform for more frequent hanging below branches, and this in turn would have selected for longer forearms, fingers, and toes, increased capacity for wrist adduction, and reorganization of the torso to enhance forelimb adduction. Knees and elbows would also have been selected to become even more mobile and support loading in a wider variety of postures.

All of these morphologies do not come as a thoroughly integrated package, as illustrated by Nacholapithecus. On the other hand, Pierolapithecus and perhaps Morotopithecus show a different suite of adaptations, with a broader rib cage and pelvis and more widely set pectoral girdle, yet with a hand and foot that displays shorter phalanges. This pattern is reminiscent of that seen in extant great apes, and to a lesser extent Oreopithecus and Dryopithecus. These variations underscore the growing realization that hominoids evolved below-branch arboreal adaptations more than once. Combining fossil data with data from extant apes, it seems increasingly likely that many of the postcranial and locomotor specializations of great apes may have evolved from ancestors that were more generalized, although not, perhaps, as generalized as basal hominoids like Proconsul.

The hypothesis that such homoplasy has possibly occurred multiple times within the Hominoidea should not necessarily be dismissed out of hand. All hominoids likely evolved from similar ancestors with a deliberate clambering mode of positional behavior. With similar selective pressures to exploit terminal branch arboreal settings, selection favored similar adaptations. Ateline monkeys
resemble hominoids in having a slightly reduced lumbar region, broader thoracic cage, longer clavicle, mobile limb joints, and other adaptations to frequent suspension. *Epipliopithecus* also resembled lesser apes in its semisuspensory adaptations (Zapfe, 1960), yet bears no close phylogenetic relation to apes.

The *Sivapithecus* dilemma initially appeared to suggest that perhaps postcranial anatomy could not be used to make phylogenetic inferences among hominoids. However, more recent discoveries of partial skeletons of *Dryopithecus*, *Equatorius*, *Nacholapithecus*, and *Pierolapithecus* show that while there indeed may be postcranial homoplasy, the variation among these taxa is greater than we had supposed. Patterns of similarities and differences can, in fact, inform our phylogenetic understanding, and so our understanding of the patterns of selection that led to the evolution of various hominoid taxa. Once we can accurately interpret the phylogenetic relations among these taxa, we can reconstruct vectors of morphological change within clades. These vectors, in turn, will provide a record of the directional selection that shaped the various lineages. This, then, provides us the opportunity to interpret the pressures that led to the observed morphological change and thus why different hominoid taxa evolved. Which is what we want to know above all.

The current picture of hominoid postcranial evolution is, therefore, one of recurring homoplasy. Different taxa represent different experiments in exploiting terminal branch arboreal habitats at relatively large body sizes, which requires the distribution of weight among different supports. Extant apes have all done this in somewhat different ways as did fossil ones. So rather than thinking about progressive stages evolving from an ancestral taxon like *Proconsul* to one like a modern great ape, we need to refocus our efforts and to explore the detailed similarities and differences among this growing collection of fossil ape skeletons to explore the diversity of locomotor adaptations within hominoids. The presence of homoplasy in hominoid evolution must not be looked at as an all-or-none phenomenon.

This approach to considering morphology is perhaps most visible to anthropologists in the origin of hominins. At present, there is no reason to assume that the most recent chimpanzee–human ancestor had such a dramatically stiffened torso, narrow cranial thorax with cranially oriented glenoid fossae and narrow scapula, or long metacarpals, and it may or may not have been a knuckle-walker when terrestrial. Instead, a slightly less derived ancestor, perhaps something like *Pierolapithecus* or *Oreopithecus* in torso morphology and perhaps other characters, might be a more reasonable guess. Only more careful morphological analysis of known fossils, comparative anatomical study of extant apes, and of course discoveries of new fossils will lead to a better resolution of the ancestral condition.
The earliest definitive hominin, and still the earliest formally described hominin taxon, *Australopithecus*, differs from later hominins in retaining morphologies that evolved as adaptations to arboreal locomotion, yet shows unequivocal strong selection for habitual terrestrial bipedality. The adaptive significance of these primitive morphologies is inherently unknowable. The extent to which *Australopithecus* and even earlier putative hominins, such as *Ardipithecus*, *Orrorin*, and *Sahelanthropus*, differed from the last common ancestor reveals the vector of selection that shaped hominin origins. For us to understand why our own lineage diverged from that of other apes, we must understand the ancestral morphology from which we evolved. Only then can we determine, for example, whether australopith forearms reduced from the primitive condition or not. The expanding hominoid fossil record affords us ever more opportunities to appreciate the diversity past and present. What may seem at first like an increasingly confusing picture of homoplasy must be viewed as an opportunity to obtain a more accurate and nuanced understanding of the adaptive diversity present throughout the hominoid radiation.

References

Begun DR (1994b) Relations among the great apes and humans: new interpretations based on the fossil great ape *Dryopithecus*. Yrbk Phys Anthropol 37: 11–63


Huxley TH (1863) Evidence as to man’s place in nature. Williams and Northgate, London


Knußmann VR (1967) Humerus rlna und radius der simiae. S. Karger, Basel
Napier J, Davis P (1959) The forelimb skeleton and associated remains of Proconsul africana. Foss Mamm Afr 16: 1–70
Molecular and Developmental evolution 302 (3): 241–267


Swindler DR, Wood CD (1973) An atlas of primate gross anatomy, Kriege, Malabar


7 Hominoid Cranial Diversity and Adaptation

Alan Bilsborough · Todd C. Rae

Abstract

The hominoid cranium represents a tightly constrained, functionally and developmentally integrated structure subject to multiple selective influences. Modern apes are the remnant of a much more diverse radiation, raising issues about their suitability as models for earlier hominoids. Among gibbons the folivorous siamang is cranially distinctive. The markedly airorynchous *Pongo* is cranially highly variable and lacks the anterior digastric muscle, thereby contrasting with other hominoids except *Khoratpithecus*. African apes share a common cranial pattern differentiated by varying growth rates, not duration. Airorhynchy is common among fossil hominoids and differentiates hominoids from non-hominoids, suggesting African ape klinorhynchy is derived. Bonobos are cranially smaller, lighter, and less dimorphic than chimpanzees. These are comparatively uniform with extensive overlap between subspecies, whereas gorillas display considerable contrasts, especially between east and west populations. Early Miocene hominoids are already cranially diverse, with most species probably soft- or hard-fruit feeders. Middle and Late Miocene forms from Africa, Europe, and western Asia are thicker enameled with more strongly constructed crania suggesting harder diets, although *Dryopithecus* (soft frugivory) and *Oreopithecus* (folivory) are exceptions. South and East Asian fossil hominoids’ diets ranged from soft fruits through harder items to bulky, fibrous vegetation. All extant ape crania are relatively lightly constructed compared with fossil forms, again prompting questions about their suitability for the adaptive modeling of earlier hominoids.

7.1 Introduction

The hominoid fossil record has expanded markedly over the last two decades, sufficiently to indicate marked morphological diversity. This is turn reflects a major radiation or, more likely, series of radiations. The great bulk of the material is from Miocene contexts—there is still comparatively little from Plio-Pleistocene deposits, apart from hominins—so that fossil and living hominoids are largely
detached from each other. Whatever the details of this array, it is clear that the extant nonhuman apes represent but the surviving fragment of a significantly more numerous, geographically more extensive, and ecologically more diverse group of catarrhine primates. The extremely restricted modern comparative base and its at best tenuous links with the earlier material pose real challenges for adaptive or phylogenetic interpretations of the fossil hominoid record.

An outcome is that detailed phylogenies often differ appreciably from author to author, depending on the significance accorded to particular apomorphies, and on the extent to which other similarities are deemed homoplasies. The upshot is a whole series of individual phylogenies and widespread disagreement about the status of particular groups which usually translate through into the taxonomies preferred by individual researchers. Since the thrust of this chapter is primarily adaptational, we do not concern ourselves with taxonomic or phylogenetic details; in what follows suprageneric categories are used informally and generally follow majority consensus usage. For those requiring more detailed information on phylogenetic issues concerning the Miocene hominoid record, the chapters by Begun, Ward, etc. in Volume 2, and the papers by Harrison (2002), Begun, Ward and Duren, Kelley, Rae (1997), and contributors, in Hartwig (2002) are excellent recent surveys.

Cranial form is influenced by multiple factors. Functionally, the head houses the visual, olfactory, and auditory organs, those of vocalization, taste, and balance; it contains the openings for the respiratory and alimentary tracts; it houses and protects the brain. It incorporates structures for food acquisition and processing, while postural and respiratory factors influence basicranial morphology. Its superficial tissues may be patterned and convey information to conspecifics about sex and ontogenetic status. The interplay of these features, and especially the size and configuration of those concerned with food processing relative to neurocranial proportions, may lead to the development of external structures such as crests and tori on the skull. There are clearly intense selection pressures determining effective developmental and functional integration of these varied aspects of cranial function throughout the individual life cycle.

Adult African apes and humans (hominids) share a broadly common pattern of covariation in cranial traits, with the oral and zygomatic regions primary integrative influences and with a lesser contribution from the nasal region, i.e., those craniofacial components primarily associated with mastication (Ackermann 2002, 2005). This differs from the pattern in both Old and New World monkeys in which the oral region is the exclusive primary contributor to facial integration. Ackermann suggests that this contrast may reflect innovatory functional or developmental shifts after the differentiation of hominoids from other Anthropoidea, or be an allometric consequence of increased body size.
Orangutans and gibbons were not represented in the analyses, but the extent to which they share the primary oral/zygomatic integrative pattern should help decide between these possibilities and assist in determining whether the pattern is a hominoid or hominid synapomorphy.

There are similar allometric patterns in the mid-face and common opposite relationships between lower and upper face in the adults. Whereas visual inspection and morphological distance place adult Pan troglodytes and Gorilla gorilla close together and Homo sapiens distant, craniofacial covariation patterns accord with molecular data in indicating closer affinity between P. troglodytes and H. sapiens, with G. gorilla distant (Ackermann 2002). However, such concordance does not hold throughout ontogeny, with differing patterns of affinity between juveniles and subadults of the above taxa on the one hand, and infants and “adolescents” on the other (Ackermann 2005).

Nonetheless, some general patterns emerge: in particular, across the species earlier and later (subadult and adult) integration appear to reflect different drivers. Oral integration is especially influential in the earlier stages as well as thereafter, but there are specific differences in the onset of zygomatic integration. In P. paniscus and P. troglodytes, it appears during the juvenile/adolescent periods, whereas in Gorilla it occurs from infancy, perhaps a correlate of its rapid growth. In all species zygomatic integration intensifies in later ontogeny. Where evident, nasal integration occurs in mid/late ontogeny, its intensity varying inversely with oral integration, suggesting that separate developmental modularities underlie these regions. While the most highly integrated species as adults, humans are more developmentally labile than the African apes prior to maturity. However, while differing in detail, all species show a common pattern of intensified integration throughout development, with a particular shift toward more constrained variation around sexual maturity or just after. The extent to which these similarities reflect, shared, genetically determined, developmental pathways or common selection pressures associated with vital functional requirements—the need for effective food processing mechanisms, for instance—remains to be determined. In the latter case, some proportion of the resemblance could be homoplastic.

While some cranial features are relatively invariant in catarrhines (e.g., positioning of orbits; structure of the auditory region), others (e.g., orbital size and shape) are highly variable within genera, species, and even subspecies (Seiffert and Kappelman 2001). Some features seem to be determined less by their “primary” function than by influences reflecting the interactions of other functional systems: e.g., the size and proportions of the orbits appear to be determined more by the growth trajectories of the mid- and upper face, and by requirements to resist the biomechanical forces generated by food processing as
they affect those regions, than by the dimensions of the visual organs housed within them (Schultz 1940). Other traits (e.g., the structure of the nasal floor and premaxilla/palatal relationships (Ward and Kimbel 1983; Ward and Pilbeam 1983)) exhibit contrasts, the functional basis of which is poorly understood but which serve as useful phylogenetic indicators (see below).

The compilation of long lists of character states as the raw data for computer-based cladistic analyses has been criticized by some (Rak 1983; Asfaw et al. 1999, 2002) as resulting in the fragmentation or “atomization” of morphology as multiple discrete traits, rather than an integrated whole. It is therefore worth noting here recent accounts that stress the importance of broader functional and developmental perspectives in analyzing morphology, and its evolutionary/phylogenetic and adaptive contexts (Lovejoy et al. 1999; Lieberman et al. 2000a, b; McCollum and Sharpe 2001; Ackermann 2002, 2005; Lovejoy et al. 2003). These build upon earlier studies such as those of Moss and Young (1960), Moss (1973), Enlow (1968, 1990), and Cheverud (1982, 1996); and biomechanical analyses such as that of Endo (1966); see also Rak (1983).

An example of this approach is McCollum’s analysis of *Paranthropus* cranial morphology (McCollum 1999; McCollum and Sharpe 2001), which concludes that limited changes in the relative growth rates of jaws and teeth on the one hand, and of the orbit and upper face on the other, would be sufficient to produce in mature individuals the distinctive set of features that characterize the robust australopithecine cranium/face. Such growth rate changes are doubtless under simple, limited genetic control and, as such, are readily evoked in appropriate selective contexts. It is not difficult to envisage comparable pressures operating on Miocene hominoids, and a variety of cranial forms thus rapidly resulting from relatively limited genetic changes. So, for example, the contrasting morphologies of *Proconsul* and *Afropithecus* might both be derived relatively simply from an Oligocene precursor such as *Aegyptopithecus*, and purely phenetic measures of affinity between these forms could be seriously awry as indicators of phylogenetic relationship.

One outcome of cladistic studies has been general recognition of the pervasiveness of homoplasy in the fossil record. From an adaptive perspective, instances of homoplasy can provide important clues as to the contexts of, and likely selective forces impacting on, hominoid communities. In such cases, the influence of phylogenetic constraint and contingency may be considerable. Minor initial differences between spatially distributed populations of a single species (or of closely related species), when further influenced by bottlenecks or other stochastic factors—easily occurring in small, localized arboreal groups, where gaps in tree cover impede gene flow—may result in significantly different morphological outcomes as responses to common selection pressures associated with
similar niches. The evolution from the nasal/palatal structure seen in Proconsul and other early Miocene forms of distinct anatomical configurations for that region in mid/late Miocene Afro-European and Asian hominoids may be an example of such a process and its outcomes.

A fundamental division of extant hominoids is that between gibbons (hylobatids) and large-bodied apes—the Asian orangutan (Pongo) and the African chimpanzee and bonobo (Pan), gorilla (Gorilla), and human (Homo), although the last taxon will be discussed elsewhere. Pongo and Pan are both largely frugivorous, with common dental adaptations (large anterior teeth and relatively small check teeth with enamel wrinkling) but differing in cranial features, whereas the more herbivorous Gorilla closely resembles Pan cranially despite its contrasting dietary niche (see below). These differing patterns of affinity illustrate the importance of developmental constraints and phylogenetic inertia in determining morphology and thus the lack of any necessary one-to-one correspondence between morphology and adaptation (for further discussion of this see below).

It is possible in principle to extend the limited insights provided by the few extant great apes into the earlier radiation by supplementing them with modeling based on early hominins, which can be thought of as phenetically and adaptively “apes” in some respects. Apart from the dangers of circular reasoning (using modern ape data as inputs into constructing early hominin models that are then used to “extend” the ape comparator base) and the appropriateness of such models (what form and degree of terrestrial orthogrady, if any, is compatible with using hominins as analogues for non-hominins?), however, there are major issues of contextual relevance.

All extant apes (here and throughout meaning non-hominin hominoids) and early hominins are essentially from tropical contexts (forest, woodland, and savannah) with none present in higher latitudes, reflecting a comparatively narrow environmental range compared with earlier ape habitats. Even incorporating early hominins within the comparator base provides a time depth of little more than 4+ Myr, characterized by broadly modern faunas that include groups rare or absent in the earlier record. In contrast, Miocene hominoids are components of markedly distinct and diverse faunas, often including entire mammalian families now extinct. So community relationships within earlier faunas will have differed from contemporary ecological webs, and the place(s) of earlier hominoids in their ecological communities are unlikely to correspond closely to those of modern ape analogues. An obvious primate example of this is the expansion and radiation of cercopithecoids over the last 10 Myr or thereabouts, so forming a major dimension of the community ecology of all recent hominoids, unlike that of earlier taxa. Floral communities also fluctuated as climatic conditions changed,
with notable contrasts between early Mid Miocene habitats and those of the later Miocene and Pliocene.

Against these differentiating features are some factors that make for modeling continuity: the range of potential (plant) food items is limited, and their physical properties even more so, limiting the nature and magnitude of the masticatory forces influencing hominoid cranial morphology. Metabolic and biomechanical constraints on body size and on locomotor form and activity, allometric influences on growth, and the functional and developmental inter-dependences of cranial form noted above all allow for a more comparative approach to hominid cranial variation. Below we review the probable ancestral condition for Hominoidea, then examine some aspects of cranial form in extant nonhuman hominoids before summarizing craniodental information on the more complete fossil forms.

### 7.2 Ancestral hominoid cranial morphology

The combination of outgroup analysis of extant forms and the morphology of stem catarrhines provides an indication of the ancestral hominoid morphotype. The Fayum fossil primates represent an early diversification of basal catarrhines, presumably reflecting dietary specialization. For example, the small and dentally and gnathically primitive *Catopithecus* (35.5–36 Ma) combines the characteristic 2.1.2.3 dental formula, postorbital closure (primarily formed from the zygomatic), fused frontals, and C1/P3 honing facet with triangular upper molars with only limited hypocone development and lower molars with high trigonids and sharp crests. *Catopithecus* had a deep and projecting face, with an especially broad premaxilla, small, widely separated orbits, and a small neurocranium with anteriorly prominent temporal lines merging to form a sagittal crest along the rear half of the vault and well-developed nuchal crests. The tympanic region is like that of platyrrhines, not catarrhines, and the mandibular symphysis is unfused. The anterior dentition displays broad, spatulate incisors and projecting, dimorphic canines, suggesting a predominantly frugivorous niche.

Many of these features, including the contribution of the premaxilla to facial proportions, small neurocranium with marked muscle attachments and pronounced ectocranial cresting, and ceboid-like tympanic region are also seen in the younger (33.1–33.4 Ma) and dentally more derived proplopithecid *Aegytopithecus*. The zygomatic is again deep and the face in general strongly constructed, with a characteristic angled profile, and the mandibular symphysis is fused. The gonial region is strongly constructed and the ramus broad and high.
The interorbital distance is again broad, with bony septa separating the high, narrow orbits and the interorbital region projects anteriorly from the medial orbital margins. Semicircular supraorbital tori extend over each orbit and, meeting medially, anteriorly bound a diamond-shaped frontal planum, whose posterior limits are defined by the anterior temporal lines. The anterior teeth are small compared to the postcanine dentition, making a narrow anterior palate. The molars are inflated and highly bunodont, especially the second, and the elongated lower third molar has a centrally placed hypoconulid; the trigonid is reduced in occlusal area and height, and lacks the paraconid, while the talonid is expanded with a large distal fovea. The upper molars are quadritubercular, with a well-developed hypocone. There is marked canine dimorphism, with upper canine honing capabilities increased by a lengthening of the anterior surface of P₃. Overall morphology points to the generation of greater occlusal pressures than in *Catopithecus*, and a craniofacial form better able to withstand the resulting forces.

When the details of stem catarrhine facial morphology are considered with the evidence from extant outgroups of the Catarrhini (e.g., Platyrhinini), it is possible to infer the major changes that underlie the ancestral hominoid craniofacial skeleton. Unlike stem catarrhines or platyrrhines, hominoids are characterized by a palate that is wide at the level of the canines, nasals that are nonprojecting and lie near the medial orbital margin in transverse section, and a premaxillomaxillary suture that contacts the nasals inferiorly near the nasal aperture (Rae 1999). Unlike previous interpretations, it is also evident that the overall shape of the ancestral hominoid morphotype is more cercopithecine-like (Benefit and McCrossin 1991), with tall zygoma and a deep face. This suggests that the shared craniofacial configuration of gibbons and colobine monkeys (short face, sloping zygoma) is convergent.

### 7.3 Extant hominoids

#### 7.3.1 *Hylobates*

Gibbons represent a radiation of small-bodied, brachiating suspensory hominoid species with attendant postcranial specializations, distinguished from each other primarily by pelage color and patterning and by vocalization. Four main groups are usually recognized, sometimes accorded subgeneric or generic rank, depending on the author. Three groups—*Hylobates hoolock*, *H. concolor* and *H. syndactylus*—are comparatively well defined; the *H. lar* group is more problematic.
Valuable reviews of extant gibbon characteristics and diversity include Groves (1972), Marshall and Sugardjito (1986), and Groves (2001); see also Geissmann (2002) and Mootnick and Groves (2005) for recent findings on gibbon diversity that support generic distinction.

Gibbons are craniodentally primitive in some characteristics (see above), compared with other extant hominoids, whether by plesiomorphy (McNulty 2004) or reversal (Rae 2004); appreciation of this led to the realization that similarities with Miocene taxa, such as *Limnopithecus* and *Pliopithecus*, previously taken as grounds for regarding these as likely gibbon ancestors, do not betoken any especially close phylogenetic relationship. The upshot is that, in the absence of a fossil record other than dental remains from Quaternary deposits of China and Indonesia referable to the modern genus, gibbon phylogeny is wholly obscure.

Overall, the *Hylobates* skull is rather lightly constructed (Figure 7.1). The neurocranium is thin walled, the vault low and ovoid in profile with a capacity of about 80–125 cm³. The frontal extends rearward between the parietals, and in most individuals the sphenoid sutures with the parietal on the vault wall. The orbits are rectangular and relatively large, with strongly developed lateral margins; a torus also develops laterally above the orbits but is not continuous, fading out medially. The lacrimal fossa extends beyond the orbital rim onto the maxilla, and the interorbital breadth is large; the short, broad nasals are usually fused above the ovoid nasal aperture. Overall the face is short, broad, and fairly projecting. Within the nasal cavity, the premaxilla and maxillary palatine process are separated by broad palatine fenestra linking the nasal and oral cavities (Figure 7.3 upper left); the vomer extends only as far as the fenestra, and the bony nasal septum is continued anteriorly by the premaxillary prevomer, which fuses to the vomer and forms a small bony crest in the incisive region in all gibbon species except the smallest, *H. klossi* (McCollum and Ward 1997). The palate and mandible are long; both corpus and symphysis are comparatively lightly built, although external thickening of the latter may be evident in some individuals, as well as the usual internal reinforcement by a superior transverse torus. The ramus is short, broad, and vertical, with some expansion of the gonial region.

Reflecting gibbons’ predominantly frugivorous niche, the anterior dental arcade is relatively broad compared with the rear. The upper incisors are markedly heterodont—I¹ broad and spatulate, I² narrow and pointed—the lowers more similar, vertically implanted and subequal in size. The canines are long, curving, transversely slightly narrowed, and sharply pointed, with minimal sexual dimorphism. There is well-developed honing of the upper canine against the long, highly compressed anterior face of the sectorial P₃, which is orientated in line with the molars. Cheek teeth exhibit considerable metric and morphological
variation, but the rear molars are usually reduced compared with the first, and especially the second molars except in H. (Symphalangus) syndactylus (see below).

The basicranium is long, with the foramen magnum and occipital condyles well behind the auditory meatus; there is no distinct mastoid process. The nuchal area is quite extensive, rising well up the occipital, with a distinct crest laterally that usually fades medially, although it may be continuous in some individuals. A sagittal crest is usually absent but may occur in small-brained individuals.

Detailed accounts of intra-and interspecific variation in *Hylobates* are given in Groves (1972, 2001) and Marshall and Sugardjito (1986), as above. Albrecht and Miller (1993) summarize their reanalysis, with caveats, of Creel and Preuschoft’s (1976) craniometric data: canonical variate analysis (CVA) reveals *H. hoolock*, *H. concolor*, and *H. syndactylus* as cranially distinct from each other.
and from the *H. lar* group. This consists of a primary cluster including *H. lar*, *H. agilis*, *H. moloch*, and *H. muelleri* subspecies, with *H. pileatus* as an outlier, and *H. l. vestitus* and *H. klossi* grouped together as a second, distinct, outlier. A subsequent analysis (Creel and Preuschoft 1984) produced patterns of resemblance that generally accord with geographical distribution but not always with the usually recognized species limits.

The only distinctive form noted here is the siamang *H. (S.) syndactylus* (Figure 7.1, lower). This large, heavily built gibbon is more folivorous than other taxa, has a larger cranial capacity, long, broad palate, and an inflatable air sac in the throat to aid calling. Postorbital constriction is more marked and, despite the larger cranial capacity, sagittal cresting is both more frequent and larger than in other gibbons, an allometric correlate of greater body size (see below).

In the dentition, the canines are less lingually curved than in other gibbons, the protocone on P³ and P⁴ larger; and on the upper molars, crowns are elongated, the hypocone variable in size, and lingual cingula almost always absent. Third molar reduction occurs in only a minority of cases and some individuals possess supernumery molars. Again consistent with its more folivorous niche, relative shearing crest development is greater than in other gibbon species (Kay and Ungar 1997, 2000).

*H. (S.) syndactylus* has a larger, more airorhynchous (i.e., more dorsally flexed) face than other gibbons (Shea 1988)—see below.

### 7.3.2 *Pongo*

The Asian great ape, the orangutan, exhibits a distinctive overall cranial form (Figure 7.2). In profile the large face is markedly prognathic subnasally, with a projecting, convex alveolar clivus. The comparatively small neurocranium is set above the facial skeleton, so that both frontal and occipital contours are relatively vertical. The orbits are elliptical, with their major axis vertical, and are surmounted by separate semicircular supraorbital costae rather than a continuous torus. The interorbital distance is very small, the ethmoid correspondingly constricted and set at a lower level than in the African apes (Shea 1988). There is no frontoethmoid sinus (Figure 7.4), and the floor of the anterior cranial fossa forms a large part of the orbital roof (Winkler et al. 1988). In the fossa, the two wings of the frontal bone fail to meet behind the ethmoid, which retains contact with the sphenoid. The nasal bones are small, typically fused at an early age, and continue beyond the frontomaxillary suture, extending as a narrow wedge into the glabellar region of the frontal. On the medial orbital wall, the lacrimal sutures with the ethmoid.
The mid-face region is short, the zygomatics are wide, deep, and flared, and there is usually a pronounced notch on the zygomatic process of the maxilla. The nasal cavity is tall and broad, the maxillary sinuses invade the interorbital pillar (sometimes as far superior as the frontal), and the lateral maxillary walls are obliquely inclined. The convex nasoalveolar clivus passes smoothly into the nasal cavity, extensively overlapping the anteriorly thin maxillary palatine process without a stepped incisive fossa; the fossa and canal are narrow, the latter long and orientated almost horizontally (**Figure 7.3** lower left). The vomer usually extends to the rear of the incisive canal but occasionally does not, in which case a small prevomer may be present (Ward and Kimbel 1983; Ward and Pilbeam 1983; McCollum and Ward 1997). Overall the palate is orientated anterosuperiorly.

The mandible is massive, the symphysis reinforced by a robust superior transverse torus and an especially pronounced inferior transverse torus extending back as far as P$_4$ or M$_1$ (Brown 1997). The corpus is deep and comparatively short. As in the African apes, there is a strongly developed platysma muscle extending laterally over much of the facial musculature and strongly attached to the swollen base of the mandibular corpus from the symphysis to the area of masseter insertion. Brown and Ward (1988) speculate that the massive platysma is associated with the orangutan’s extensive laryngeal air sac system—greater than in other apes—aiding the regulation of air pressure and volume within the sac during vocalization. A distinctive feature of *Pongo* is absence of the anterior
digastric muscle (and so of the digastric fossae on the base of the symphysis), and associated separation of the posterior digastric from the hyoid and stylohyoid muscle (Dean 1984; Brown and Ward 1988). Instead the large posterior digastric, originating on the cranial base adjacent to \textit{rectus capitis lateralis}, inserts onto the gonial region between the medial pterygoid and masseter muscles, acting to depress the mandible. The orangutan’s mylohyoid muscle is especially well developed, as are the geniohyoids. \textit{Rectus capitis lateralis}, originating from a narrow area on the front of the atlas and inserting on the basioccipital anterior to the foramen magnum, is a more fan-shaped muscle than in the chimpanzee.

The cranial base is wider than in the African apes (Dean and Wood 1981, 1984), but the eustachian process much smaller, providing the origin for only \textit{tensor palati}, with \textit{levator palati} originating from the apex of the petrous temporal (Dean 1985). The mastoid processes are poorly developed. In the articular region there is a long preglenoid plane, an indistinct articular eminence, and a prominent postglenoid tubercle. The roof of the glenoid fossa is coronally oblique, slightly sloping inferomedially, so that the entoglenoid is less prominent than in the African apes. The temporomandibular ligament is well developed.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure7_3.png}
\caption{Subnasal morphology of hominoids seen in sagittal section. Upper left: \textit{Morotopithecus}, showing no overlap of the premaxilla on the maxilla (the primitive condition seen in extant \textit{Hylobates} and most fossil hominoids); Lower left: \textit{Pongo}, showing the smooth overlapped subnasal condition also seen in \textit{Sivapithecus}. Right: \textit{Pan} (upper) and \textit{Gorilla} (lower) showing the stepped overlapped condition usual in extant African apes. Modified after Ward and Kimbel (1983)}
\end{figure}
laterally but lacking the deeper horizontal band, suggesting closer approximation of the rear of the working condyle and the postglenoid tubercle during chewing (Aiello and Dean 1990).

The foramen magnum and occipital condyles are set well back on the skull base. A nuchal crest is present in all mature individuals, and a prominent sagittal crest develops posteriorly in most males, uniting with the nuchal crest but, reflecting the orangutan’s greater airorhynchy, typically not extending as far beyond the rear of the vault proper as in Gorilla (see below). Anteriorly the temporal muscles diverge as lines or simple crests bounding a triangular area of the frontal. As in the African apes, the bulk of the temporalis muscle is orientated obliquely, with an emphasis on the posterior fibers.

The dentition reflects the orangutan’s predominantly frugivorous niche. The upper incisors are the most heteromorphic of any extant hominoid: $I^1$ is very broad and spatulate, but $I^2$ is smaller, more pointed and more convex in curvature. Well-developed median and marginal ridges reinforce the incisor crowns in biting. Lower incisors, high crowned and narrower than the uppers, are also reinforced by lingual ridging. Canines are conical, markedly dimorphic and especially robust in males; females display more pronounced lingual cingula. Upper premolars are bicuspid; $P_3$ is sectorial with a narrow, elongate protoconid as the honing face; $P_4$ is bicuspid. Upper molars are more oval in occlusal outline than in other apes. Cheek teeth are relatively large compared to body size, low crowned and with extensive, deep secondary wrinkling that further increases occlusal area. Molar shearing crests are rather well developed considering the emphasis on fruit (although significant quantities of bark and leaves are also ingested), exceeding those of chimpanzee species but considerably less than gorillas (Kay and Ungar 1997). They perhaps provide an instance of phylogenetic inertia, suggesting a more folivorous ancestor.

Orangutans are remarkably variable in cranial morphology (Wood Jones 1929; Röhrer-Ertl 1988a, b). Röhrer-Ertl (1988b) has shown that the most stable region is the mid-face, other cranial areas varying according to age, sex, dental eruption and masticatory development, hormonal status, dietary composition, and tooth use. Both neurocranium and face exhibit greater growth in breadth than in length or height, a differential that is more marked in males than in females. While there is much individual and intrapopulational diversity, at least some variation reflects geographic factors: Groves (1986, 2001) and Röhrer-Ertl (1988a, b) review cranial patterning and Brown (1997) mandibular form, while Uchida (1998b) summarizes dental differences. Within a context of admittedly high variability, Sumatran orangutans are characterized by an oblique but straight (not concave) facial profile with highly protuberant anterior teeth, a
convex cheek region lacking a suborbital fossa, relatively short nasals, a shorter neurocranium but with a longer nuchal region, and a longer foramen magnum. The mandibular symphysis tends to be long and narrow, with an extensive inferior transverse torus. Dentally they exhibit relatively small paracones on P³ and M¹ compared with their Bornean counterparts, M¹ larger than M² rather than subequal, and a broader M₃.

Bornean orangutans have a generally more prognathous and concave facial profile, display a distinct suborbital fossa on the cheek, have more labially positioned incisors, a “trumpet-shaped” nasal aperture that becomes triangular in cross-section at the level of the nasal tubercle (Röhrer-Ertl 1988a), and a more prominent interorbital pillar (Groves 2001). Their mandibles are deeper and broader anteriorly, and the symphysis is usually larger, thicker, and more bulbous than that of Sumatran orangutans. There is also craniodental differentiation within Borneo between populations from Sabah, NW and SW Kalimantan separated by the Kapuas River (Groves 1986, 2001; Courtney et al. 1988; Groves et al. 1992), often of comparable magnitude to that between Bornean and Sumatran orangutans.

For example, Uchida (1998b) was unable to identify any consistent pattern of dental differences between Pongo populations from W Borneo, SW Borneo, and Sumatra, with the Bornean groups often as distinct from each other as either was from the Sumatran sample. Bornean orangutans were significantly different from each other (but not from Sumatra) in P₄ and M₁ shape, but virtually identical in their narrow M₃ shape, with Sumatran orangutans having broader rear molars. Differences in molar cusp proportions showed similarly inconsistent patterning between the three groups. There were no obvious links to dietary differences, and Uchida concluded that on dental evidence, river and mountain systems within Borneo were as significant biogeographic barriers, and so promoters of differentiation, as flooding of the Sunda shelf.

Bornean and Sumatran orangutans have generally been accorded subspecific status as Pongo pygmaeus pygmaeus and P. p. abelii, respectively. In his latest revision, however, Groves (2001) distinguishes them as separate species (P. pygmaeus and P. abelii) on the basis of the more comprehensive morphological information now available, and molecular differences well above levels usually associated with subspecies, which indicate a long period (c. 1.5 Ma) of isolation between the two forms. He also formalizes the intra-Bornean diversity noted above as subspecies of P. pygmaeus.

Hominoids exhibit more dorsal flexing of the face relative to the cranial base (airorhynchy) than non-hominoids; their orbital axes and palates are both shifted more dorsally relative to their degree of basicranial flexion than those of other primates (Ross and Ravosa 1993; Ross and Henneberg 1995). While the functional basis for this is disputed (Ross and Ravosa 1993) and may well have
multiple causes, within this context many of the orangutan’s distinctive features can be plausibly related to its extreme airorhynchy (Delattre and Fenart 1956, 1960; Biegert 1964; Shea 1985, 1988; Brown and Ward 1988). Biegert (1964) argued that the hypertrophied laryngeal sac in Pongo is a prime determinant of its skull form, comparing it with the enlarged hyoid and associated throat organs of Alouatta. Shea (1985, 1988) and Brown and Ward (1988) have criticized this interpretation. Shea considers laryngeal specialization as just one potential determinant of airorhynchy, interacting with other factors, largely unknown. Brown and Ward consider the Pongo–Alouatta analogy invalid in view of contrasts in the submandibular anatomy of these two genera, and it has also been rejected by Hershkovitz (1970) and Zingeser (1973).

Shea argues that pronounced dorsal flexion of the face links Sivapithecus and Pongo, and that a degree of airorhynchy (although not to the extent seen in these two genera) is primitive for catarrhines and hominoids generally. On this view, the more ventral positioning of the face relative to the neurocranium seen in African apes and hominids is synapomorphous and, as such, a significant phylogenetic indicator (see also Ross and Ravosa 1993; Ross and Henneberg 1995; and below).

### 7.4 The African apes

As is well known, the African apes (here and throughout meaning gorillas, chimpanzees, and bonobos, i.e., non-hominin hominines) share a basic similarity of cranial form, and in many respects are scaled variants of a common bauplan (Figures 7.5 and 7.6). Many of the craniodental differences between them have been related, with varying degrees of success, to differences in dietary niche (see Chapters 8, 11–15). Taylor (2002) also provides a useful recent summary of African ape diets. Within a highly variable context of local preferences and seasonal fluctuations, and with considerable overlap in the fruits exploited, gorillas are, broadly speaking, more folivorous than chimpanzees. Gorillas consume less fruit than chimpanzees and exploit leaves, pith, bark, bamboo, and terrestrial herbaceous items. The Eastern Mountain gorilla (G. g. berengei) is the most severely folivorous form; the western lowland gorilla (G. g. gorilla) exploits the most varied diet, with a significant fruit component. In contrast, chimpanzee diets are dominated by fruits, although it is unclear whether the bonobo (P. paniscus) exploits more terrestrial herbaceous vegetation than the common chimpanzee (P. troglodytes) (Taylor 2002).

Compared with the orangutan, African apes exhibit longer, lower, narrower neurocrania set at a lower level relative to the facial skeleton (klinorhynchy).
The frontal contour is low and retreating, the parietal region flat, and the occipital more curved than in the large-bodied Asian ape. There is a prominent supraorbital torus that is usually continuous across the glabellar region as well as above each orbit, although in some *P. troglodytes* individuals it may be divided by a slight depression. A supratoral sulcus, its lateral limits defined by the anterior temporal lines, delimits the torus from the frontal squama. The orbits are subrectangular, usually broader than high, and interorbital breadth is greater than in the orangutan, reflecting the broader ethmoid of African apes. On the medial orbital wall, the ethmoid’s orbital plate is reduced and the ethmolarcral suture is usually much less extensive than in the orangutan, and in some individuals may be replaced by contact of the interposed frontal and maxilla. There is an extensive frontoethmoid sinus (**Figure 7.4**). On the floor of the anterior cranial fossa, the frontal may separate the ethmoid from the sphenoid, more commonly in *Gorilla* (>50%) than *Pan* (15%). Frontotemporal contact predominates on the lateral cranial wall of the chimpanzee and gorilla, but sphenoparietal contact is common in the bonobo.

The root of the maxillary zygomatic process arises relatively close to the occlusal plane, above M1 or M2. In the chimpanzee, the zygoma’s facial (malar) aspect is limited in height and breadth; in the gorilla it is deeper and extends further laterally. In both apes it is remarkably thin in sagittal cross section when compared with most early hominins but is strengthened by the sagittal angulation.
of its upper and lower portions. Rak (1983) has emphasized the structural importance of the zygomatic region as a transverse buttress, linking the lateral and medial components of the face, and resisting masticatory forces. In both gorilla and chimpanzee, the zygoma’s temporal process is sharply angled from its malar surface, with the zygomatic arches orientated parasagittally/posteriorly slightly divergent (Pan) and parasagittaly/posteriorly slightly convergent (Gorilla), reflecting differing ratios of mid-facial and bitemporal breadths in the two genera. The greater facial breadth in Gorilla means that the masseters, especially their anterior fibers, have a greater lateral component to their contraction than in Pan.

The zygomatic arch is thin in cross section but vertically deeper, its inferior border marked anteriorly for the superficial masseter fibers, and in Gorilla posteriorly scalloped for the origin of the muscle’s deeper portion. A part of this, sometimes differentiated as the zygomaticomandibularis muscle, fuses with anterior temporalis fibers, to attach to the temporalis tendon, the coronoid process, and anterior ramus edge (Raven 1950; Sakka 1984; Aiello and Dean 1990). In mature male gorillas, the arch is reinforced sagittally in its mid-region by a “step” with convex upper border which increases its vertical depth compared with immediately adjacent areas, and strengthened transversely toward its rear by the broad, flat base of the temporal’s zygomatic process. Additional support against the masseter’s pull is provided by the temporalis fascia, inserting on the upper border of the zygomatic arch; again, it is particularly extensive in male gorillas.

Anteriorly the face is braced against masseteric force by the zygomatic buttress (see above) and by the beam of the supraorbital torus, which links with the zygoma via its frontal process (Rak 1983). The greater facial breadth of Gorilla combined with its more marked postorbital constriction and so deeper infratemporal fossa means that the lateral component of the torus is unsupported behind by the anterior neurocranial wall, and so is massively thickened vertically and sagittally, while the postorbital bar is broadened compared with Pan. These structures, the canine roots, and nasal septum also reinforce the palate and face against bending (sagittal), torsional (coronal), and shearing forces generated during biting by the anterior teeth. Such forces are highest rostrally and of greatest magnitude in large-jawed forms such as Gorilla (Preuschoft et al. 1986).

Within the nasal cavity, the incisive canal is wide and the fossae are broad and bowl shaped. In Pan, the extent by which the premaxilla overlaps the palate, and so the length of the incisive canal, is comparable to that in the orangutan, although the canal is angled more steeply than in the latter because of the African ape’s less convex premaxilla (Figure 7.3 upper right). In Gorilla, the overlap is much less and the incisive canal shorter, and there is always a distinct step in the
nasal floor between premaxilla and palate (Figure 7.3 lower right). In Pan, the step is much less marked and may be absent altogether in about one third of individuals, who evince a smooth floor comparable to that of the orangutan (McCollum and Ward 1997). In Gorilla, a long prevomer is interposed between the vomer and the premaxilla, with the inferior parts of both the former bones descending into the incisive canal, dividing its posterior wall and eventually partitioning it into two channels. A septal groove along the nasal sill is seen only in younger individuals; in adults it is confined to the rear of the sill. In Pan, the prevomer is much smaller and, while it descends into the incisive canal to divide the posterior wall, together with the vomer, complete partitioning into two channels is much less frequent than in Gorilla. Unlike the latter, a septal groove is present on the nasal sill in adults as well as younger individuals.

Fusion of the facial aspect of the premaxillomaxillary suture in chimpanzees begins prenatally and is usually completed before the permanent dentition is fully erupted, with the nasal aspect being completely fused around the eruption of M2. Facial growth in Gorilla continues for longer, with both facial and nasal aspects of the premaxillomaxillary suture and the prevomer–vomeral sutures remaining open until well into maturity (McCollum and Ward 1997). Accessory premaxillary sutures are also quite common (>20%) in Gorilla, indicating separate ossification centers for the palate and facial components of the premaxilla (Schultz 1950).

The palate is long in both Pan and Gorilla; externally it is shallow anteriorly with no clear alveolar border but deeper along the postcanine row. Internally, the maxillary palatine process of Pan is distinctive in thickening anteriorly and containing the palatine recess, a medial extension of the maxillary sinus. Laterally the maxillary alveolar process is thin, with the contours of the tooth roots evident; medially the process is thicker. Rak (1983) argues that the maxillary zygomatic process acts as a mid-palatal buttress, reinforcing the hard palate against shearing stresses generated between the chewing and balancing sides of the dental arcade, primarily from the latter’s medial pterygoid muscle. Both medial and lateral pterygoids are particularly well developed in Gorilla.

As in the orangutan, the preglenoid plane of African apes is long, the articular eminence only slightly developed so the glenoid fossa is sagittally shallow, and the postglenoid tubercle is well developed. The roof of the glenoid fossa is coronally more horizontal than in the orangutan and the entoglenoid more distinctly differentiated from it, especially in Gorilla, where it is very large, extending beyond the level of the articular eminence and preventing any medial shift of the condyle prior to moving onto the preglenoid plane (Du Bruhl 1977). In some of these features, and in temporal bone shape overall, Pan is more
derived than *Gorilla* (Lockwood et al. 2004). A prominent temporomandibular ligament is present in *Gorilla* and is apparently variably developed in *Pan* (Aiello and Dean 1990).

In *Pan* species the dentitions are basically similar, although *P. paniscus* teeth are smaller and less sexually dimorphic than those of *P. troglodytes*. The maxillary incisors are curved mesiodistally, with I₁ larger than I₂, although the difference is smaller in *P. paniscus* than in *P. troglodytes*. In the mandibular incisors, these proportions are usually reversed. The upper canine is larger in males than females of both species; its mesial surface is more convex in *P. troglodytes* and, with the lingual surface, displays grooving absent in a small sample of *P. paniscus* (Swindler 1976). In the upper jaw, M₁ and M₂ are subequal in size, M₃ reduced, with the hypocone the smallest cusp and reducing progressively along the molar row. Reduction is more pronounced in *P. paniscus*, and the cusp may even be completely absent from M₁ and M₂ in some individuals, whereas it is always present on those teeth in *P. troglodytes*. The hypocone may be entirely absent on some M₃s of both species but is more weakly developed in bonobos (fully developed in 21% of *troglydotes* teeth, compared with only 9% of *P. paniscus*). The preprotocrista (anterior transverse crest between paracone and protocone) is more angled and transversely orientated in *P. paniscus*, running from closer to the protocone to mesial of the paracone rather than to its tip, as in *P. troglodytes*. The distoconule, an accessory cusp between hypocone and metacone, is absent in bonobos but present in all chimpanzee subspecies, generally at low frequency but up to 40% of M₃ in one collection of *P. t. troglodytes* (Kinzey 1984). A lingual cingulum is often present, most frequently on M₁ but larger on M₃ and better developed (longer distally) in bonobos than chimpanzees.

M₂ is usually the largest mandibular molar, M₃ the smallest; a Y-5 cusp pattern is almost universal on M₁ but only occurs in <50% of cases on M₃. The talonid is extensive, and a buccal cingulum is rarely (5%–10%) present (Swindler 1976). In *P. paniscus*, the metaconid is usually opposite the protoconid rather than distal to it as in *P. troglodytes*, resulting in a greater relative distance, and a deeper groove, between the metaconid and entoconid in the former species (Kinzey 1984). Nonetheless, the two cusps are closely adjacent compared with *Gorilla*. The hypoconulid is usually slightly buccally positioned in chimpanzees, and more centrally (lingually) placed in bonobos, while a tuberculum sextum is often present between hypoconulid and entoconid in the former species but more rarely in the latter, which Kinzey (1984) suggests may be associated with the more lingually positioned hypoconulid. *Pan* molars are often wrinkled but not to the extent seen in *Pongo*. 
Central to lateral incisor proportions in Gorilla are comparable to those of Pan, although compared to the postcanine teeth, the incisors are much smaller. Canines are large and markedly dimorphic, in the female projecting less beyond the other teeth. Contrary to the sequence in Pan, but like the orangutan, P₄ erupts before P₃, which is sectorial but with a vestigial metaconid, a large distal fossa for the P₃ protocone, and a well-developed lingual cingulum. On the upper molars, the hypocone is larger relative to the other cusps than in other apes; the mesial fossa is narrow, the distal one wide, and a lingual cingulum is usually present. On the lower molars, the metaconid and entoconid are widely separated and there is an extensive talonid basin to receive the large protocone of the upper molar. A tuberculum intermedium is often present between metaconid and entoconid on M₁ and is almost invariably so on M₂ and M₃; a tuberculum sextum may also occur between the entoconid and the buccally positioned hypoconulid. A buccal cingulum is usually present on M₁, on about 50% of M₂, and on a minority of M₃; overall it is both more common and better developed in Gorilla than in other extant apes. In the upper jaw, M² is usually the largest tooth; in the lower jaw, M₁ is the smallest, with M₂ and M₃ subequal (Swindler 1976). Dimorphism in dental dimensions is extensive in Gorilla, with most teeth differing significantly in size between the sexes. Tooth enamel is smooth, without the wrinkling displayed by Pongo and Pan. Supernumery molars may occur, more often in the upper jaw than the mandible.

Temporal muscles are well marked on the cranial walls in the chimpanzee, often forming raised ridges which in mature males may occasionally meet to form a sagittal crest. In male gorillas, a pronounced sagittal crest is present, thickened at the top where the two temporal laminae abut, and highest toward the rear of the vault where it unites with the nuchal crest, forming a beaklike posterior projection at the rear of the skull. The crest, besides enlarging the area for temporalis attachment, improves the power of the cheek teeth by increasing the relative length of the muscle insertion axis compared with the load, and also serves to increase the effective height of the neurocranium, thereby enhancing its resistance to the vertical forces generated during mastication (Davis 1964).

A compound T/N crest (Robinson 1958) forms laterally in chimpanzees from the juxtaposition of the temporal and nuchal muscles, but these diverge medially, and there the perimeters of the temporal and nuchal muscles are marked by lines, slightly raised ridges, or a simple nuchal crest. In male gorillas, the nuchal muscles develop medially as well as laterally beyond the neurocranium proper, producing a compound T/N crest uniting with the sagittal crest as above, and resulting in an extensive, triangular-shaped nuchal area.

Temporalis fibers originate from the lower part of the nuchal crest’s anterior surface but do not attain its rim, which provides attachment for the
occipitofrontalis scalp muscle (Sakka 1984; Aiello and Dean 1990). Medially trapezius and laterally sternocleidomastoideus insert on the posterior rim of the nuchal crest, with below these the rhomboids (medially), and the fleshy, laterally extensive splenius capitis muscles. Deep to these is the heavy semispinalis capitis, which may be divisible into medial, thick biventer, and more lateral, straplike, complexus portions (Aiello and Dean 1990), although this separation is said to be uncommon in Pan (Swindler and Wood 1973) and is not indicated in Raven’s (1950) account of Gorilla anatomy.

On the cranial base, rectus capitis lateralis lies immediately lateral to the mid-rear portion of the occipital condyle in Gorilla and Pongo, and to the front-mid-portion of the condyle in Pan (Dean 1984; Raven 1950). It is unclear whether the rather more anterior insertion of the muscle in the chimpanzee reflects individual variation or a specific trait. Just lateral and slightly posterior to this muscle is the digastric; see above for its distinctive form in the orangutan. Just in front of the foramen magnum and close to the midline are the closely adjacent rectus capitis anterior muscles, and ventral to these the longus capitis muscles.

7.5 Individual species morphology and intraspecific diversity

7.5.1 Pan

7.5.1.1 Pan troglodytes

The commonly recognized subspecies may be distinguished cranially as below, based primarily on Groves (2001).

P. t. troglodytes possesses a very broad head combined with a comparatively narrow muzzle, a continuous, straight, medially thickened supraorbital torus, more concave facial profile, and more gradually sloping occipital than other subspecies. On the medial orbital wall, ethmo-lacrimal contact is very common, while supernumerary bones on the lambdoid suture are rare, as are multiple infraorbital foramina.

P. t. verus also has a broad, rather flat-topped cranium but a broader muzzle, a less sharply concave facial profile, and a more steeply curved occipital. The supraorbital torus is arched over each orbit and is laterally well developed. Ethmo-lacrimal contact is very rare, while additional bones at lambda and along the lambdoid suture are very common. The frequency of a single infraorbital foramen bilaterally is higher than in other subspecies.
P. t. schweinfurthii has a more rounded skull than other subspecies, with an elongated, gently sloping occipital. The facial profile tends to be straight or only slightly concave, and the muzzle narrow, although interorbital breadth is high. The straight, continuous supraorbital torus is thinner than in other subspecies, especially laterally, but is prominent at glabella. Multiple infraorbital foramina are very common, and frontotemporal contact at pterion virtually universal. In cranial nonmetric traits generally it resembles P. t. troglodytes but is rather smaller and less sexually dimorphic than that subspecies. Despite this, Angst (quoted in Groves et al. 1992) has reported a higher average cranial capacity for schweinfurthii—420 cm³, compared with virtually identical capacities for troglodytes and verus (401 and 404 cm³, respectively). Highly variable in size and cranial proportions, schweinfurthii may incorporate more than one subspecies.

P. t. vellerosus is a recently recognized subspecies from Nigeria to Cameroon (Gonder et al. 1997), identified on mtDNA sequencing that showed it to be a sister taxon of P. t. verus. Cranially it is unlike P. t. verus but similar to P. t. troglodytes and P. t. schweinfurthii in its high frequency of ethmo-lacrimal contact and low frequency of Wormian bones at lambda and along the lambdoid suture (Groves 2001).

7.5.1.2 Pan paniscus

Bonobos are characterized by smaller heads and teeth than common chimpanzees but comparably sized upper limbs, rather lighter, more slender trunks, and heavier hindlimbs (Zihlman 1984). The bonobo skull is smaller, smoother, and more lightly built than that of the chimpanzee, the mandible appreciably shorter, the face considerably less prognathic, and reduced in height (Figure 7.5). Reflecting the less projecting face and jaws, the cranial base is more tightly flexed, with a mean angle of 140° compared with 145° in the chimpanzee (Cramer 1977). This flexion results from a basicranial growth pattern to adulthood in P. paniscus that resembles that of P. troglodytes curtailed at the subadult (M² eruption) stage (Laitman and Heimbuch 1984)—see also below. The bonobo supraorbital torus is thinner and the supratoral sulcus weaker, while the frontal squama rises (and the occipital descends) more steeply than is usual in P. troglodytes. It is more common (57%) for the sphenoid and parietal to suture at pterion (contrast P. troglodytes above), while on the orbital wall frontomaxillary contact is more frequent than in chimpanzees (24% and 9%, respectively, Cramer 1977).

While bonobos exhibit some canine dimorphism, there are only very limited differences between sexes in the size of the incisors and cheek teeth (see above).
Similarly, mean endocranial capacity is virtually identical in males and females at c. 350 cm³ compared with 404 and 375 cm³, respectively, in *troglodytes* (Cramer 1977). The nuchal area may be bounded by a low ridge or line, but a true crest with sharply defined rim is absent, as is any sign of sagittal cresting. Consistent with its more neotenous form, *P. paniscus* shows earlier closure of the facial component of the premaxillary/maxillary suture than *P. troglodytes* and much higher frequency of a completely open palatal component (>93% cf. 19%, respectively, of individuals with M¹ erupted) (Braga 1998). This early synostosis results in a vertically and horizontally shorter face and reduced dental arch, consistent with the bonobo’s significantly smaller incisors, compared with *P. troglodytes*. Kinzey (1984) notes the greater degree of incisor wear in *P. paniscus* than *P. troglodytes*, which he suggests may be related to a greater incidence of pith
and leaf petioles in the diet; he also speculates that the combination of a more transversely orientated and angled preprotocrista, with a more mesially sited metaconid and deeper groove between protoconid and hypoconid into which the crest occludes (see above) produces a more efficient shearing mechanism that again may reflect a more folivorous dietary component in bonobos.

Morphometric studies illustrate the relative homogeneity of chimpanzee cranial form compared with other great apes. While usually distinguishing *P. paniscus* from *P. troglodytes*, differentiation within the latter is, not unexpectedly, less secure, with extensive overlap between subspecies; see, for example, Shea and Coolidge (1988). These authors found that discrimination just about reached the subspecies threshold, and that separation was considerably less than in orangutans or gorillas (see below). They considered that this comparative uniformity might reflect a more recent differentiation of *P. troglodytes* subspecies, more frequent or extensive contact—and so gene flow—between them, marked ecological flexibility for the species overall so precluding close matching of subspecific features to habitat, or any combination of these. A subsequent study (Groves et al. 1992), with specimens sorted by location rather than subspecies, produced neither meaningful geographic patterning nor subspecific grouping among males. However, female crania exhibited better separation, with *P. paniscus* distinct, *P. t. schweinfurthii* grading geographically toward *P. t. troglodytes*, and with evidence for east–west differentiation within *P. t. schweinfurthii* based on facial proportions.

Shea et al. (1993) compare the results of both raw and size-adjusted analyses. For the former, there is 100% correct classification for *P. paniscus* females and about 75% correct classification for *P. troglodytes*, of which *P. t. verus* and *P. t. schweinfurthii* are furthest apart, according with their geographic separation. However, confining the analysis to *P. troglodytes* removes this geographic gradient, with maximal separation now between *P. t. troglodytes* and *schweinfurthii*. As expected, size adjustment reduces separation of *P. paniscus* from *P. troglodytes* so that the distance between *P. t. verus* and *P. t. schweinfurthii*, now the most widely divergent subspecies, approaches that between the latter and *P. paniscus*. Principal Components Analysis shows *P. paniscus* clustering with immature *P. troglodytes* crania along PC 1, indicating their common growth trajectories and emphasizing that shape contrasts between bonobo and chimpanzee reflect the smaller size and truncated growth of the former relative to the latter, within which the major differences between *P. t. troglodytes* and *P. t. schweinfurthii* are also due to size and associated allometric factors (see below).

Separate analysis of mandibular variation in *Pan* accords generally, but not completely, with the above (Taylor and Groves 2003). Mandibular separation within *P. troglodytes* is less than that within *Gorilla*, but contrasts between
P. paniscus and P. troglodytes are greater than Gorilla, and there is clear separation of bonobos and chimpanzees. There is extensive overlap of P. troglodytes subspecies, maximally between P. t. schweinfurthii and P. t. troglodytes, and greatest distinction between the latter and P. t. verus (contrast to Shea et al.’s cranial finding of greatest overlap between P. t. troglodytes and P. t. verus). Size adjustment again reduces separation, so that bonobos, while remaining the most distinctive, now partly overlap with chimpanzees; and P. t. verus, while still the most isolated of chimpanzee subspecies, is now furthest from P. t. schweinfurthii (as on the cranial data). P. t. verus’ distinctiveness on mandibular traits, while relatively slight (Taylor and Groves 2003), nonetheless accords with Braga’s finding (1998) that premaxillomaxillary suture closure differs significantly between P. t. verus and other subspecies, with P. t. verus displaying later complete closure of the suture’s facial component and earlier closure of its palatal component compared with P. t. troglodytes and P. t. schweinfurthii (Braga 1998). This points to a longer, deeper lower face in P. t. verus than other subspecies.

7.5.1.3 Gorilla

Most accounts of Gorilla cranial diversity are based on Groves’ highly influential morphometric analysis of variation in 45 traits from >700 gorilla skulls, grouped by origin into 19 and 10 geographic localities for crania and mandibles respectively (Groves 1967, 1970). D² values were calculated for each of ten cranial and six mandibular representative variables, allowing the localities to be grouped into eight larger regions which could be further combined on the basis of intra- and intergroup differences into three clusters: a relatively homogeneous western cluster (four regions, of which the Cross River sample was rather more distant from the other three), a distinctive eastern group from the Virunga volcano region, and a further eastern group (three regions). These correspond to the western lowland gorilla (G. g. gorilla), the eastern highland gorilla (G. g. beringei), and the eastern lowland gorilla (G. g. graueri) (Figure 7.6).

G. g. gorilla is the smallest subspecies, with fairly broad face, small jaws and teeth, a short palate, a single mental foramen under P₃ or P₄ (more usually under the latter), and a jaw condyle without a cleft. G. g. graueri is the largest subspecies, with a high, narrow face, larger jaws and teeth, and a longer palate. The mental foramen is often multiple and set under P₃, while the jaw condyle is often cleft. The mountain gorilla, G. g. beringei, is distinguished by a low, broad face, very large jaws and teeth, a very long tooth row and palate, anteriorly sited (under C or P₃) multiple mental foramina, and a jaw condyle that is usually cleft.
Stumpf et al. (1998), adjusting for size, demonstrated that the Cross River sample was more distinctive than Groves’ original analyses indicated, so providing support to the growing movement advocating its recognition as a further subspecies, *G. g. diehli* (see Sarmiento and Oates 1999, 2000; Groves 2001, 2003). *G. g. diehli* is distinguished by its shorter skull, shorter molar row, narrower palate, shorter cranial base, and more steeply angled nuchal plane than other western gorillas, which Sarmiento and Oates speculate may be associated with a diet of smaller, drier, and harder food items than that of other western gorillas.

A further reanalysis of Groves’ data (Stumpf et al. 2003) confirmed a primary east–west separation on the latter’s smaller values for palatal and tooth row lengths, nasal aperture and nasal bone breadths, lateral facial height, and supraorbital torus thickness. They also demonstrated the distinctiveness of the Cross River and
Virunga populations from other west and eastern groups on the basis of their narrower interorbital breadths, narrower palates, and reduced lateral facial height. Analyses restricted to the western populations further indicate the distinctiveness of *G. g. diehli* on overall and neurocranial lengths, bicanine and bimolar breadths, interorbital and neurocranial widths, palatal length, and medial and lateral facial heights. However, Stumpf et al. emphasize that the fundamental distinction is between east and west *Gorilla* populations, with the corollary that *G. g. graueri* is more closely related to *G. g. beringei* than it is to western lowland gorillas. The implications of this, together with recent data from molecular and other studies, have led Groves (2001, 2004) to revise his earlier taxonomy, and to differentiate western and eastern gorillas at the species level as *G. gorilla* (*G. g. gorilla* and *G. g. diehli*) and *G. beringei* (*G. b. beringei* and *G. b. graueri*), respectively. This also accords with the zoogeographical evidence, but for consistency with other sources referred to herein we retain the traditional single species classification.

Leigh et al. (2003), however, apply Wright’s $F_{ST}$ (an indicator of microdifferentiation, measuring the extent to which subdivision within species—i.e., between subspecies—departs from random mating) to Groves’ craniometric
data and to discrete trait variation and reach different conclusions. Their approach requires assumptions about population sizes and the heritability of craniometric traits but is considered to be robust, especially, when subspecies sizes differ markedly, as they do in *Gorilla*. $F_{ST}$ calculated from the craniometric data yields unexpectedly low levels of between group variation: only c. 20% between subspecies compared with 80% within subspecies assuming equal population sizes. Adjusting for different population estimates between the subspecies results in even lower values of $F_{ST}$, with correspondingly more variation within subspecies. $F_{ST}$ derived from discrete trait analysis gives rather higher, but still modest, levels of divergence. Leigh et al. argue that much gorilla variation reflects ontogenetic changes and sexual dimorphism and, as such is intra-subspecific, and that their results offer no support for differentiating eastern and western gorillas at the specific level. See also Albrecht et al. (2003) for a detailed analysis of *Gorilla* cranial diversity at locality, deme, subspecies and species levels, and its potential evolutionary and sociobiological implications. Interestingly, genetic data indicate much deeper levels of differentiation among African ape species than the morphological evidence does (Gagneux et al. 1999; Lockwood et al. 2004).

Dental variation in *Gorilla* is considerable: molar shapes and cusp proportions, relatively invariant within subspecies, differ between subspecies, as do tooth dimensions (Uchida 1998). Male (but not female) *G. g. beringei* canines are larger than those of *G. g. gorilla* and *G. g. graueri*. In the postcanine dentition, *G. g. graueri* is, surprisingly, significantly larger than *G. g. beringei*, which is larger than *G. g. gorilla*. While upper molars of *beringei* show B-L enlargement, those of *graueri* are expanded in both length and breadth. *G. g. gorilla* has wider incisors relative to molar length than the eastern subspecies, while *G. g. beringei* displays higher crowned cheek teeth with sharper cusps and ridges than *G. g. gorilla*. *G. g. graueri* has a relatively smaller talonid on P4 and, together with *G. g. beringei*, has larger distal cusps on the upper and lower molars than *G. g. gorilla*. Patterns of dental sexual dimorphism differ among gorilla subspecies: *G. g. beringei* displays greatest dimorphism in canine and lower molar size, *graueri* greatest dimorphism in upper molars, with *G. g. gorilla* least dimorphic both in canines and molars. This reflects a larger canine relative to molar size in females of this subspecies, which Uchida considers to reflect heightened female–female competition, possibly related to greater frugivory. She stresses the importance of local dietary adaptation influencing tooth form and proportions, with considerable variation but with more extensive frugivory in *G. g. gorilla* and lowland *G. g. graueri* than in highland populations of that subspecies and *G. g. beringei*. Further analyses of the gorilla masticatory system are discussed below.
7.6 Allometric and biomechanical studies

The greater size of the gorilla relative to the chimpanzee is an instance of peramorphosis (Shea 1983c). Length of maturation is comparable in all three African ape species, but gorillas grow much more rapidly and to greater sizes than chimpanzees, while bonobos grow somewhat more slowly than chimpanzees (rate hypermorphosis) to rather lesser sizes (although there is considerable overlap); within each species, males grow for longer than females (time hypermorphosis). The pattern indicates that the interspecies differences reflect selection for greater body size, perhaps associated with increasing terrestrial folivory rather than selection for delayed maturation (Shea 1983c).

Significant differences between *Pan* and *Gorilla* growth in body weight only become apparent after about 2 years, with gorillas pulling away increasingly strongly from the chimpanzee growth curve thereafter (Shea 1983d; Figure 7.6). Since neural growth is predominantly prenatal/immediately postnatal, there is no corresponding increase in *Gorilla* brain size in later ontogeny; the natural consequence is that while having absolutely larger brains than chimpanzees, gorillas have lower brain: body ratios and lower encephalization quotients, with male value particularly depressed compared with females. “In the case of the African pongid species, . . . the developmental pathway utilised to increase body size ensures that relative brain size decreases as a consequence” (Shea 1983d p 58). It follows that attempted explanations of behavioral and/or ecological contrasts between the species based on differences in relative brain size should be regarded with skepticism.

Shea (1983a, 1984) also summarizes evidence that the differences in body form between adult *P. troglodytes* and *P. paniscus* result from ontogenetic scaling. The extension of common growth allometries to different end sizes holds within the skull, trunk, and limbs but not between them, so that adult bonobos do not match any single stage in chimpanzee ontogeny. Relative to the latter, the *P. paniscus* skull is most strongly reduced in size, forelimbs and trunk less reduced, and hindlimbs not reduced at all, so that for a given body size, *P. paniscus* has a smaller skull than *P. troglodytes*. This in turn results in a more paedomorphic cranial shape compared with the chimpanzee through the decoupling of growth rates for head and body, with the former slowed relative to the latter—an instance of neoteny. The selective factors underlying this process are obscure, although Shea speculates that the reduced sexual dimorphism and different social organization of *P. paniscus* compared with *P. troglodytes* may be important drivers in the evolution of its distinctive cranial proportions.

Earlier suggestions (Ackermann and Krovitz 2002) of a common cranial postnatal ontogenetic shape trajectory or of separate but parallel shape
trajectories that merely accentuate differences established in early (prenatal) ontogeny have been refuted by Cobb and O’Higgins (2004) who show hominin postnatal shape trajectories to be divergent with differing shape changes between species, even in early postnatal ontogeny (Vidarsdottir and Cobb 2004). However, the directions of scaling trajectories between Pan species are not significant (so changing postnatal facial shape in a similar manner from different starting points), whereas those between Pan and Gorilla are directionally distinct.

In general, the degree of adult cranial sexual dimorphism is greater in the larger apes (gorilla and orangutan) than in the chimpanzee and bonobo. Gorilla males display more size and shape variability than females, and a similar difference appears to be present in Pongo, but not in Pan (O’Higgins and Dryden 1993). Most of the differences reflect greater male facial prognathism, in turn a consequence of canine dimorphism. Adult cranial dimorphism appears to result from distinct mechanisms in the African and Asian apes. Whereas in Pan and Gorilla cranial dimorphism follows from extending the growth period of males for most cranial proportions (Shea 1983c), in Pongo only about half the growth allometries exhibit this process, with the other half displaying accelerated growth in males compared with females (Leutnegger and Masterson 1989a, b).

Male and female chimpanzees display significant cranial size differences but no shape differences, perhaps because the period of extended growth is a short one and/or the scaling coefficients are minor, so resulting in insignificant shape differences given the comparatively modest size of chimpanzees. As with the bonobo–chimpanzee comparison above, these differing patterns of cranial dimorphism in the great ape genera have been linked to socioecological contrasts between them (O’Higgins and Dryden 1993).

Shea has also explored allometric influences on African ape craniofacial and dental form and their relationships to diet using bivariate and multivariate techniques. Many facial proportions in the bonobo, chimpanzee, and gorilla exhibit ontogenetic scaling, i.e., a common pattern of size/shape change. There are also instances, however, where this does not obtain: for example, chimpanzees have shallower zygomatic roots, narrower bizygomatic breadths, smaller infratemporal fossae, and narrower anterior cranial bases than bonobos with the same basicranial lengths. In other words, these features are reduced in chimpanzees compared with the values expected in bonobos ontogenetically scaled to their sizes (Shea 1984). Similarly, in those features in which chimpanzees are reduced relative to bonobos, gorillas tend to be reduced relative to chimpanzees (Shea 1984). As Shea points out, such allometrically adjusted analyses point to the opposite conclusion from that usually drawn from the study of absolute skull
sizes—cranially bonobos are relatively the most robust, and gorillas relatively the most gracile of the African apes.

Additionally, gorillas have significantly longer and higher cranial vaults, higher orbits, and longer foramen magnums than chimpanzee crania of equivalent basicranial lengths. They also exhibit longer, more projecting nasal regions that are sited lower on the face, than comparably sized chimpanzees. As overall skull size increases in the sequence bonobo–chimpanzee–gorilla, the three species also exhibit relatively narrower faces and neurocrania, reflecting in the latter case, the fact that increased brain size results primarily from growth in length, not width, during the prenatal and early postnatal phases.

During late postnatal growth, occipital length and breadth in gorillas increase appreciably compared with chimpanzees, reflecting the development of sagittal and nuchal crests as a functional response to the enlarged temporal and nuchal muscles “outgrowing” their areas of attachment on the exterior cranial wall. This, in turn, is a consequence of the respectively positive and negative allometric relationships between splanchnocranial and neurocranial proportions and body size. Dental metrics indicate that gorillas have relatively smaller incisors and relatively larger cheek teeth than comparably sized chimpanzees, while, surprisingly, temporal fossa area (and so temporal muscle size) becomes relatively smaller across the three species as size increases. Despite these differences, the predominant pattern among the African apes is essentially one of similarity in craniofacial growth. Multivariate analysis yields a common allometry vector incorporating >93% of total variance confirming this general picture, with a second vector (3.4%) distinguishing chimpanzees and gorillas.

Shea considers that the differences in mid-face proportions between Pan and Gorilla may reflect differences in soft-tissue function or dietary contrasts, although the influence of the latter is by no means clear. In fact, he notes that while dental contrasts between the two apes can fairly clearly be linked to diet, no significant reorganization of the face occurs, with its form being primarily determined by the endpoints of common allometric trajectories. This suggests that the face and masticatory apparatus may be less strongly coupled to diet than is the dentition, and/or that chimpanzee and gorilla diets, while differing in their constituent items, may not differ appreciably in their physical properties, in particular the force required to process them. For Shea, the African ape masticatory complex provides an example of an integrated functional system preadapted to extension into new size ranges and dietary shifts.

Molar crown area scales positively in hominoids, so that larger forms have relatively as well as absolutely larger crown areas, associated with their generally increased folivory, primarily achieved through increased tooth lengths rather
than breadths (Demes et al. 1986). There are associated increases in palatal and mandibular lengths and relative narrowing of upper and lower dental arcades in larger taxa. Allometrically determined snout elongation produces greater bite force in larger animals, by lengthening the horizontal distance between the mandibular joint and the molar row, which Demes et al. have shown scales with a mean value of c. +1.6 in hominoids. Bite force is maintained by lengthening the masticatory muscles’ power arms, by increasing their cross-sectional area, or by a combination of these. The temporal muscle’s power arm scales from +1.16 (male great apes) to +1.62 (female gibbons); that of the masseters and medial pterygoids, which is strongly influenced by facial height, between +1.54 and +1.64. A rough estimate of temporalis cross-sectional area scales from c. +1.4 to +1.9, indicating that the greater load of the allometrically lengthened lever arm is more than matched in larger species by the positive allometry of the power arm and of muscle cross-sectional area, and so muscle force.

Larger species produce more bite force for their size than smaller ones due to the allometric changes in masticatory biomechanics following from increased body size, a point neatly illustrated by Demes et al. who demonstrate that skulls of H. klossi and P. paniscus enlarged isometrically to the size of H. (S.) symphylanus and G. gorilla, respectively, produce lesser bite forces than the “real” latter two forms do. Bite pressure (bite force/crown area) is maintained if bite force increases at the same rate as crown area; broadly similar relationships for these variables hold within hylobatids and great apes, indicating that in hominoids, crown area and bite force increase at about the same rate, at least over the size range of extant taxa. Estimated bite pressure is generally greater in great apes than hylobatids, although the orangutan is an exception here. Bite pressure shows no obvious relationship to between-species differences in size or diet; interestingly, P. troglodytes males produce the greatest scaled bite pressure, exceeding even G. gorilla males. Within species, males generally produce greater pressures than females, although whether this is selected for (implying differences in food processing or paramasticatory activity between the sexes) or is a by-product of selection for larger body size is a moot point. Demes et al. make the important point that similar allometric relationships obtaining within hylobatids and great apes provide strong evidence that biomechanical constraints associated with increased body size elicit similar functional responses across the Hominoida. This is strong presumptive evidence that they should also be applicable to fossil forms.

Ravosa (2000) undertook such a combined allometric analysis of mandible size and form in fossil and extant apes, comparing them with cercopithecoids. Deeper corpora counter parasagittal bending while the more robust cross sections of larger species, especially the fossils, counter axial torsion. The positive
allometry of corpus and symphysis cross sections suggests increased masticatory stresses due to greater balancing side muscle activity during powerful mastication, probably reflecting a tougher, harder diet. In addition, the allometry of jaw length and breadth point to greater wishboning stresses at the symphysis at the end of the masticatory powerstroke, countered by a thicker symphysis and increased anterior jaw breadth. After allometric scaling, the most robust mandibles include *Proconsul africanus* and *P. nyanzae*, *Rangwapithecus*, *Turkanapithecus*, *Afropithecus*, *Ankarapithecus*, *Lufengpithecus*, and *Ouranopithecus*; the more slender include hylobatids, *Simiolus*, *Dryopithecus laietanus*, *Pan paniscus*, and *P. troglodytes* (see also below).

In the broader context of fossil as well as extant hominoids, even *G. gorilla* has a comparatively low and slender corpus, only average symphyseal height and a slightly broader than expected symphysis for its mandible length. *Pongo* has a higher but rather thinner corpus than expected and a higher symphysis of expected width; *P. paniscus* has a corpus and symphysis of expected height but rather thinner than expected, while *P. troglodytes* has a shallower, narrower corpus and symphysis, although the latter is closer to the values expected on jaw length than the former. These findings have implications for the dietary reconstruction of fossil forms: many Middle/Later Miocene hominoids are notably more robust than modern apes, and so are plausibly reconstructed as exploiting harder, more resistant food items requiring substantial force in processing, while the categorization of proconsulids as “frugivorous” may also well underplay the variety and toughness of their dietary items. Such reconstruction, however, is difficult to reconcile with the evidence provided by dental morphology and wear patterns in *Proconsul* and by some other aspects of jaw form (see below).

In some respects, hominoid and cercopithecoid mandibular cross-sectional scaling patterns are similar; smaller apes are notably gracile, resembling cercopithecine proportions, but larger ones have both deeper and relatively wider corpora more reminiscent of colobines to resist greater axial torsion during chewing—perhaps reflecting larger, more laterally placed masseters that contribute relatively more to unilateral mastication, together with the medial pterygoids. This is especially so in the largest apes which have absolutely and relatively very thick corpora exceeding colobine proportions, suggesting diets with at least comparable, and very possibly greater, physical properties of hardness and/or toughness.

Smaller apes also display symphyseal curvature comparable to cercopithecines whereas, with increasing body size, curvature reduces to a shallower, colobine-like, arc, eventually falling below even that in the largest apes. This has traditionally been interpreted in dietary terms (frugivores requiring large incisors and so a wide anterior dental arcade) but the bulk of the fossil evidence points
to diets other than frugivory. Ravosa therefore interprets the broader anterior dental arcade combined with a relatively thick symphysis as hominoid adaptations to resist concentrated wishboning forces at elevated levels resulting from a hard object diet. The exception to this is *Afropithecus* which has a notably narrow, tightly curved anterior mandible with, presumably, correspondingly concentrated wishboning forces, raising interesting issues about dietary composition and food processing activities in that genus. Again, findings from the dental evidence do not easily accord with those based on mandibular proportions.

More recently, Taylor (2002, 2003) has used morphometric methods to investigate masticatory variation in African apes as a function of dietary differences. She compared allometrically adjusted mandibular, cheek and facial dimensions quantifying masticatory parameters associated with bite force and load resistance in ontogenetic series of bonobo, chimpanzee, and gorilla to test the hypothesis that more foliverous forms would show greater development of these features.

The results are complex. Unsurprisingly, all species show allometric increases during growth in traits indicating improved muscle and bite force and the capacity to resist greater loads. Masticatory muscle sizes are especially strongly allometric. However, after adjusting for allometry only a few traits differ consistently across African ape species as predicted by dietary preferences. A more resistant diet is generally correlated with a thicker mandibular corpus, although the thicker corpus of chimpanzees compared with bonobos is not matched by evidence of corresponding dietary differences. Compared with *Pan*, *Gorilla* has a relatively wider mandibular corpus to resist axial torsion, a wider symphysis so resisting “wishboning,” a higher temporomandibular joint which contributes to improved mechanical advantage of the jaw lever and distributes forces more evenly along the cheek teeth, and a higher mandibular ramus, increasing the moment arm of the temporal and masseter muscles, and providing a larger attachment area for the latter and the medial pterygoids. Moreover, within *Gorilla*, eastern gorillas exhibit greater values than western ones and also have larger masster muscle than the latter, in accord with their more resistant diet.

Other analyses, however, do not conform to the pattern predicted from diet—for example, gorillas do not have the relatively deeper corpora expected to resist parasagittal bending (Hylander 1979a, b); there was no regular association of the deeper symphysis providing increased resistance to bending and shearing forces with greater folivory (see also Ravosa’s findings summarized above, although Demes et al. (1984) and Wolff (1984) considered the torsional resistance of the gorilla mandible “remarkable,” and maximal at the symphysis). Gorillas overall do not have the shorter deeper faces and more anteriorly
positioned masticatory muscles predicted to improve the mandible’s power arm ratio and to reduce bending moments in the face, and there was no consistent differentiation of bonobos and chimpanzees. Taylor concludes that while some of the distinctive craniofacial features of the African apes can plausibly be considered as dietary adaptations, the link is not especially strong. Dental development and allometric and other ontogenetic constraints are doubtless important influences, while more information is needed on the composition, variability, and especially the physical properties of ape diets. The equivocal nature of these results accord with those from some other studies: for example, despite their dietary contrasts, Rak’s indices quantifying relationships between the palate, masseter origin, and their positions relative to the calvaria fail to discriminate between *Pan* and *Gorilla* (Rak 1983: table 3 p 25).

Further analysis of the *Gorilla* masticatory system with larger samples and including *G. g. graueri* in the analysis confirm and extend the earlier findings (Taylor 2003). *G. g. beringei* has a significantly larger face than the other two subspecies, and differs from *graueri* in the same features that distinguish it from *G. g. gorilla*, and which differentiate *Gorilla* and *Pan*. However, despite being more folivorous, *graueri* does not differ from *G. g. gorilla* in those features, and *G. g. beringei* fails to express the full set of masticatory traits predicted by its diet. In this last respect, it may well be the case that an investigative model assuming optimization of each and every variable is simply inappropriate: rather than a spectrum or continuum of values for every trait, some may be more appropriately considered in terms of thresholds. For example, if food availability is not a constraint, and provided the face is structurally sufficiently strong to resist masticatory forces, there seems little reason to suppose that selection will necessarily promote further shortening and deepening of the face in hominoid folivores to achieve “optimal” values, particularly if to do so will disrupt pervasive, well-established allometric trajectories. Covariation within a tightly integrated functional system, such as the head, is likely to impose multiple constraints on the variation of any given character or character complex.

Given this lack of concordance in masticatory morphology and diet in living hominoids, and the incongruities between craniofacial morphology and dental evidence among fossil forms, detailed dietary reconstruction based on craniofacial form in fossil hominoids appears questionable. There seems no secure basis on which to go beyond the most general of statements about dietary properties. Shea’s conclusions about the relative decoupling of masticatory morphology and diet in the African apes may be applicable to hominoids generally. Dental evidence (crown proportions and morphology, crest development and structure, chemical composition, and wear patterns) may well prove a better guide to hominoid diets than the analysis of craniofacial form, no matter how elaborate
the biomechanical models employed. These provisos should be kept in mind in relation to the following summary of craniofacial form in fossil hominoids, which for its framework draws especially on the chapters by Harrison, Begun, Kelley, and Ward and Duren in Hartwig (2002).

### 7.7 Fossil hominoids

#### 7.7.1 Africa

#### 7.7.1.1 Early Miocene/Oligocene African hominoids

**Kamoyapithecus** The earliest possible hominoid is Kamoyapithecus, from Late Oligocene deposits at Lothidok, Kenya. More extensive evidence from Early Miocene sites in East Africa indicates an array of forms, variously regarded as hominoid or non-hominoid (see Begun Volume 2 Chapter 4). Evidence from the oldest (19–20 Ma) sites of Songhor and Koru (Kenya) points to predominantly tropical forest habitats, with later (16–17 Ma) sites on Rusinga Island (Kenya) ranging from flood plain to riverine contexts. The picture here is of drier, more seasonal environments than Songhor or Koru, but with persistent wooded conditions, varying from forest to deciduous woodland according to rainfall (Andrews 1996; Andrews et al. 1997). Similar habitats are indicated at the rather later Middle Miocene sites of Maboko Island (15–16 Ma) and Fort Ternan (14.5 Ma) (see below).

**Proconsul** At least two broad cranial morphologies are represented among the Lower Miocene fossils. The genus Proconsul is particularly well known, with species differing in size, dental and gnathic details, but linked by fundamental similarities in craniodental morphology (Figure 7.7). Knowledge of Proconsul is primarily based on the material recovered from sites on Rusinga Island, Kenya, including much of a skull in 1948 (Le Gros Clark and Leakey 1951), and a partial skeleton from 1951 and subsequently (Napier and Davis 1959; Walker and Teaford 1989; Walker et al. 1993) supplemented by other material. Initially assigned to the type species *P. africanus*, the 1948 and 1951 specimens were later transferred to *P. heseloni* (Walker et al. 1993) on the basis of differences in dental size and morphology and in mandibular proportions and symphyseal reinforcement from the type of *P. africanus* from Koru.

Compared with later hominoids, the *P. heseloni* cranium is lightly constructed: the globular neurocranium largely lacks pronounced tori or crests, although the medial portion of the nuchal crest is evident above the steeply
angled nuchal area, and the external occipital protuberance is located high on the skull rear. The frontal is short but broad, reflecting the limited postorbital constriction, while the superior temporal lines are prominent anteriorly and converge toward the vault rear but do not meet to form a crest. On the face, the premaxilla rises on either side of the nasal aperture, contacting the nasal bones above their tip, so excluding the maxilla from the rim of the nasal aperture, which narrows inferiorly between the central incisor roots above a short nasoalveolar clivus. Above it the nasal bones are nonprojecting, long, and narrow, extending upward beyond the frontomaxillary suture and expanding in breadth in the glabellar region. There is a prominent jugum above the upper canine root and a shallow canine fossa. The lightly built zygomatic arch originates low down, curving backward and upward, and has a well-developed malar tuberosity. The subrectangular orbits are widely separated, surmounted by weak supraorbital
ridges and with a slightly swollen glabella over a large frontal sinus (which may represent a frontoethmoid sinus; see Rossi 2005) between, but there is no distinct supraorbital torus.

The palate is long, rectangular, and shallow. A large, transversely broad incisive fossa joins directly with the nasal cavity, so there is no true incisive canal, and the hard palate is retracted from the subnasal alveolar process (Ward and Pilbeam 1983; McCollum and Ward 1997). There is a pronounced tuberosity on the alveolar process behind $M^3$. The maxillary sinus extends anteriorly to the premolars and laterally into the root of the zygomatic arch. The articular eminence and postglenoid process are well developed, and the auditory region has a tubular ectotympanic as in modern catarrhines, while the prominent, well-pneumatized mastoid process is coronally narrow and rather bladelike.

The mandibular symphysis exhibits a moderate- to well-developed superior transverse torus but a much weaker and variable inferior torus, absent altogether in some individuals. The corpus displays limited lateral buttressing below the cheek teeth and shallows posteriorly; the relatively high ramus is lightly constructed, and the gonial region only slightly marked by muscle attachments. The slightly earlier, similarly sized and evidently closely related $P. africanus$ differs from $P. heseloni$ in numerous dental details, mandibular proportions, and symphyseal reinforcement: the $P. africanus$ corpus is deeper anteriorly and posteriorly shallows more strongly, while the symphysis lacks an inferior transverse torus but bears a pronounced superior torus (Walker et al. 1993). An inferior transverse torus is also absent in the larger $P. nyanzae$ and $P. major$.

Views differ on Proconsul’s brain size and encephalization: Walker et al. estimated the 1948 cranium’s capacity as 167 cm$^3$ and so inferred that $P. heseloni$ was more encephalized than modern cercopithecoids of similar size. Manser and Harrison (1999), however, using foramen magnum area as a surrogate estimated brain size as the markedly lower 130 cm$^3$ and close to the mean encephalization value for anthropoids. The endocast’s relatively small frontal lobe and cortical sulcal pattern were considered primitive and cercopithecid by Le Gros Clark (Le Gros Clark and Leakey 1951; Le Gros Clark 1962) but definitely not so by Radinsky (1974) who judged it hominoid and most like gibbons. More recently, Falk (1983) has argued that the sulcal pattern resembles that of extant New World monkeys, such as Ateles, rather than any group of catarrhines and approximates the inferred common ancestral sulcal pattern for Anthropoidea.

Relative shearing crest lengths in those Proconsul species studied ($P. heseloni$, $P. nyanzae$, $P. major$) are less than those of any extant ape and considered to indicate frugivory (Kay and Ungar 1997), while the proportion of pits to scratches on molar wear facets (37%–39%) also indicates soft-fruit eating (Ungar
et al. 2004). Despite the findings of Ravosa (2000), see above, mandibular form and proportions are also compatible with this interpretation. At the symphysis, the prominent superior transverse torus and minimal or absent inferior torus produce a broadly triangular cross section, especially in larger individuals, resistant to torsional or bending stresses produced by medial (jaw opening) or lateral (masticatory power stroke) bending (Hylander 1984; Brown 1997). Below the cheek teeth, the corpus is relatively deep vertically and narrow coronally and, while it may be reinforced by the rear of the superior transverse torus, the mylohyoid ridge, and the lateral eminence, there is relatively little change in cross-sectional shape along the corpus compared with other fossil apes (Brown 1997). This suggests chewing activity generating comparatively high vertical forces but only limited transverse or torsional forces during food processing.

Afropithecus The contemporary Afropithecus turkanensis contrasts markedly with Proconsul in its cranial form (Leakey and Walker 1997) (Figure 7.8). The large face is dominated by the long domed muzzle and deep, flaring zygomatic processes. The projecting premaxilla forms a deep nasoalveolar clivus and extends up on both sides of the broad, oval nasal aperture to contact the narrow, medially elevated nasal bones. There is an extensive maxillary sinus; the shallow palate displays large paired openings for the incisive foramen. The canine roots form prominent juga; the root of the zygomatic process is deep, anteriorly inferiorly sloping, and originates low down on the face. The cordiform orbits are broader than high, inferolaterally sloping, and widely separated. The glabellar region is prominent, the slender supraorbital torus curving above each orbit, and delimiting a frontal trigon with its lateral limits marked by well-defined anterior temporal lines which merge to form a distinct sagittal crest. Postorbital constriction is marked, and the temporalis muscles are well developed. The mandibular corpus is very deep with a distinct fossa, while the ascending ramus is set obliquely to the corpus. The symphysis bears moderate superior and inferior transverse tori, a long and strongly sloping subincisive planum, and a low-genial pit, implying a deep, narrow tongue.

Overall facial proportions of Afropithecus are reminiscent of A. zeuxis but absolutely much larger; finite element scaling analysis reveals marked size contrasts but minimal shape differences in the snout, and some shape differences (but reduced size contrasts) in the zygomatic and maxillary tuberosity regions (Leakey et al. 1991). These authors conclude that the similarities in Aegyptopithecus and Afropithecus facial form indicate the persistence into the Early Miocene of a functionally integrated mosaic of features that characterized the primitive
hominoid face. Benefit and McCrossin (1991) draw attention to craniofacial similarities between *Aegyptopithecus*, *Afropithecus*, and the Miocene cercopithecoid *Victoriapithecus*, indicating that many of these facial traits are primitive catarrhine characters rather than basal hominoid synapomorphies.

With its large, procumbent and mesially inclined upper central incisors, stout, low-crowned canines and cheek teeth covered with very thick enamel and complex wrinkling, *Afropithecus* has been compared, especially in its anterior dentition, to pithecines exploiting seeds and hard fruits, where the incisors crop food items and the large canines apply considerable force to puncture hard fruits, as in *Chiroptes* (Leakey and Walker 1997; Kinzey 1992). Cusp morphology, the high incidence of pitting in the single *Afropithecus* individual sampled—at 43%, the highest of the early African forms studied (Ungar et al. 2004)—and a lack of prominent shearing crests on the cheek teeth are also consistent, like the anterior dentition, with frugivory.
Overall, Leakey and Walker (1997) conclude that *Afropithecus* was a sclerocarp forager. Other aspects of mandible reinforcement (e.g., strong basal buttressing, a pronounced lateral tubercle where the oblique line meets the corpus, a hollowed buccal surface above the mental foramen with a marked canine jugum anteriorly), which support this interpretation are also seen in *Sivapithecus* (see below), and probably reflect comparable biomechanical responses to reliance on food items with similar physical properties rather than any especially close phylogenetic link. The somewhat smaller but otherwise similar *Heliopithecus* from the early Middle Miocene of Saudi Arabia may be no more than specifically distinct from *Afropithecus* (Andrews et al. 1978; Andrews and Martin 1987).

**Morotopithecus**  The large (chimpanzee sized) hominoid *Morotopithecus bishopi* (Gebo et al. 1997), based on the palate from Moroto initially assigned to *P. major* (Pilbeam 1969), resembles *Afropithecus* in many respects and may be congeneric with it. It is of Early (20–21 Ma) or Middle (15–17 Ma) Miocene age, depending on $^{40}$Ar/$^{39}$Ar dating (Gebo et al. 1997) or faunal correlation (Pickford et al. 1999). It combines an anteriorly broad palate with comparatively narrow, procumbent incisors offset by a pronounced diastema from the large, stout canines, whose massive roots form pronounced juga. The molars resemble those of *P. major* in their bunodont cusps, wrinkled enamel and beaded lingual cingulum, but contrast in their relative sizes, while the anterior dentition is much larger and the interobital breadth narrower than in *P. major*. Overall the face is relatively long and narrow, with a broad nasal aperture, a short clivus, and an extensive maxillary sinus. The undoubted resemblances in face and dentition to *Afropithecus* may reflect dietary convergence in exploiting hard-cased fruits rather than phylogenetic propinquity. Postcrania referred to *Morotopithecus* resemble those of *Proconsul* in some respects but are also markedly more derived in the direction of modern hominoids in the lumbar, shoulder, hip, femur, and knee regions, and point to forelimb suspension and slow brachiation, as well as climbing and quadrupedal activity in an arboreal habitat (MacLatchy 2004).

**Nyanzapithecus**  *Nyanzapithecus* is much less well known, but premaxillary and maxillary fragments of two species, *N. vancouveringorum* from the early Miocene (17–18.5 Ma) and *N. pickfordi* from the Middle Miocene (15–16 Ma), indicate contrasts with *Afropithecus* in their smaller size, shorter faces, low, broad nasal apertures, and robust premaxillary regions (Harrison 2002). These species and the rather smaller Middle Miocene *N. harrisoni* (13–15 Ma) display broad, strongly built upper and lower incisors, while the cheek teeth are long and narrow; the molars bear low, expanded cusps and rounded occlusal crests.
Mabkopithecus, represented by two isolated rear lower molars and an almost complete mandible is dentally very similar to N. pickfordi, and may well be congeneric with Nyanzapithecus—in which case the former genus has priority.

Rangwapithecus The Early Miocene (19–20 Ma) Rangwapithecus gordoni, similar in size and probable locomotor pattern to the smaller Proconsul species, contrasts dentally with them in numerous respects, including molars with low cusps and well-developed crests, enamel wrinkling, and a pronounced wear differential. Cranial material indicates a comparatively short premaxilla and long, narrow palate widening toward the rear. The maxillary sinus is deep and the zygomatic root set low down the face above M1–M2; the mandible is deep and the symphysis reinforced with a pronounced superior transverse torus. Kay and Ungar (1997) and Ungar et al. (2004) have argued that on the basis of its molar crest development (greater than that of any other Early Miocene form) and low incidence and long, narrow form of dental pitting, Rangwapithecus is likely to have been folivorous. Dental proportions and macrowear, facial morphology, palatal proportions, and mandible structure are all compatible with this interpretation.

Turkanapithecus The somewhat younger (16.6–17.7 Ma) Turkanapithecus kala-kolensis is another medium-sized form, rather smaller than P. heseloni, and represented by a partial cranium preserving the upper dentition save for the incisors, and a mandible with left M2 and right M3 (Figure 7.8). The skull exhibits a relatively short face with broad, domed snout, a wide, oval nasal aperture flanked by prominent canine pillars, with expanded nasal bones and a broad, flat interorbital region above. The palate is narrow, with posteriorly convergent toothrows, and there is an extensive maxillary sinus. The zygomatic process originates low down on the face, and the arch is relatively deep and flaring. This, combined with pronounced postorbital constriction, makes for a deep infratemporal fossa and, presumably, well-developed temporalis muscle—a view also supported by the strongly marked and convergent temporal lines, pointing to a sagittal crest. The rear of the saddle-shaped glenoid cavity is bounded by a well-developed postglenoid process, but there is no distinct articular eminence. The nuchal area is comparatively short and the crest strongly developed, reflecting both the rugged facial architecture and comparatively small neurocranium, estimated at c. 85 cm³—absolutely and relatively smaller than P. heseloni (Manser and Harrison 1999).

Given this cranial morphology, it is rather surprising that the mandibular symphysis displays neither strongly developed superior nor inferior transverse tori. The corpus is shallow and relatively slender, with constant depth below the
molars, while the ramus is broad, low, and sloping, with an expanded gonial region—so according with zygomatic architecture indicating well-developed masseter muscles—and a knoblike condyle. Upper first premolars are large and while both upper and lower molar teeth increase in size posteriorly, the gradient is much less than in comparably sized Proconsul species. Overall, craniodental features suggest a resistant diet, possibly consisting of hard-cased fruits or leaves.

**Dendropithecus** There is also a cluster of small- to medium-sized forms (perhaps 3–9 kg), usually grouped together as dendropithecoïds. The siamang-sized *Dendropithecus macinnesi*, based on material originally assigned to *Limnopithecus*, displays narrow, high-crowned incisors, strongly dimorphic canines, broad premolars, and molars with high cusps, sharp occlusal crests, and well-defined foveae. The palate is narrow, as is the nasal aperture; the maxillary sinus is extensive and the mandible corpus low and robust, with the symphysis reinforced by fairly prominent superior and inferior transverse tori. Among Early Miocene hominoids, *Dendropithecus* has the least well-developed molar shearing crests of those studied other than *Proconsul*, and a fairly high incidence of pitting, pointing to a predominantly soft-fruited dietary niche (Kay and Ungar 1997; Ungar et al. 2004).

**Micropithecus** The smaller *Micropithecus* contrasts dentally in having relatively larger anterior teeth compared to the cheek teeth, broad incisors, and narrow premolars and molars, with only weakly developed occlusal ridges. It differs cranially from *Dendropithecus* in its shallow, broader palate and nasal aperture, short face and clivus, moderately high and lightly built mandible corpus, with only modest symphyseal tori. The frequency of dental pitting and pit shape in *Micropithecus clarki* (19–20 Ma) point to folivory (Ungar et al. 2004), while *M. leakeyorum* (15–16 Ma) shows similarities to the rather earlier *Simiolus enjessi* (16.5–18 Ma) that probably also reflect folivorous adaptations (Harrison 1989; Benefit 1991).

**Limnopithecus** Other small Early Miocene forms, known principally from isolated teeth and part jaws, include *Limnopithecus legetet* and *L. evansi*. These have a short lower face, with anteriorly positioned orbits, narrow, elliptical nasal aperture, and a shallow clivus; an inflated maxillary sinus, and a shallow, lightly built mandible reinforced symphyseally by a strongly developed superior transverse torus but with an inferior torus that is weak (*L. evansi*) or absent (*L. legetet*). *L. legetet* combines broad, low-crowned incisors and small canines with an ovoid P₃ that suggests only part development of the C–P₃ honing complex, and cheek teeth with high, sharp cusps and occlusal crests. *L. evansi* has narrower, higher
crowned incisors, larger canines, and a better developed sectorial face on $P_3$, cheek teeth with lower, rounded cusps, and less sharply developed occlusal crests. Relative shearing crest development suggests a fairly folivorous niche (Kay and Ungar 1997).

**Kalepithecus**  Less well known than these forms is the broadly contemporary *Kalepithecus songhorensis*, similar in dental size to *L. legetet*, but differing in most other respects. The anterior teeth are relatively large, with $I_1$ broader and more spatulate, the upper premolars relatively narrow but the molars relatively broad; $P_3$ is moderately sectorial, the short, broad lower molars with low, rounded, and expanded cusps, and poorly developed occlusal crests. *Kalepithecus* contrasts with other Early Miocene forms in its inferiorly broad nasal aperture and deep clivus, while dental morphology and proportions suggest a frugivorous diet.

### 7.7.1.2 Middle and Late Miocene African hominoids

Fossil hominins apart, evidence of African hominoids from the Middle Miocene onward is limited, especially, when contrasted with the comparatively abundant Lower Miocene material. Nonetheless, recent discoveries have both significantly increased the number of fossils and led to major reappraisals of earlier finds, notably of the material assigned to “*Kenyapithecus*” *africanus* (Leakey 1967), which has been reallocated to distinct taxa, with consequent systematic and phyletic implications.

**Nacholapithecus**  One such taxon is the large, markedly dimorphic *Nacholapithecus kerioi* (Ishida et al. 1999), based on material from Middle Miocene (15 Ma) sites in Samburu District, Kenya. Cranial remains display overlap of the posterior premaxilla and hard palate, and the incisive foramen is small, while the face bears strong canine pillars and deep fossae, and the zygomatic process originates low down on the maxilla. The mandible corpus is tall but thin, with a near vertical symphysis, a moderate inferior transverse torus, and with a lateral fossa below the premolars. $I_1$ is high crowned and robust, while both lower incisors are tall and narrow. Canines are low crowned, upper premolars display marked cusp heteromorphy, and molars are thick enameled.

**Equatorius**  Other material previously assigned to “*K.* africanus” has been incorporated, along with new discoveries including a part skeleton (KNM-TH 28860) from Kipsaramon, Tugen Hills, and multiple finds at Maboko Island and
adjacent localities, Kenya, in another, broadly contemporary (14–15.5 Ma) large-bodied, dimorphic species, *Equatorius africanus* (Ward et al. 1999). While initially criticized—see, for example, Begun (2000) and Benefit and McCrossin (2000) and response by Kelley et al. (2000)—the current consensus is that *E. africanus* is a valid taxon. The species is characterized by broad I₁s with marginal ridges, markedly asymmetrical lateral incisors, with a spiral lingual cingulum, and relatively large upper premolars with reduced cusp heteromorphy (*contra* *Nacho-lapithecus*). The procumbent, narrow lower incisors are tall whereas the mandibular canines are low crowned with convergent roots. The thick-enamed, bunodont lower molars increase markedly in size along the tooth row.

The maxilla exhibits a very low, broad root for the zygomatic process and an extensive sinus extending into the premolar region, while the mandible displays a long, inclined sublingual planum, a prominent inferior transverse torus, and a robust corpus. The partial skeleton and other postcranial fossils indicate some resemblances to earlier forms such as *Proconsul* and *Afropithecus*, but with forelimb and hindlimb contrasts that point to significant terrestriality (Sherwood et al. 2002). It remains to be determined whether ground vegetation formed an appreciable component of *Equatorius’* diet, but dental similarities to *Afropithecus* (and *Heliopithecus*), especially in canine form and premolar proportions, suggest resistant foods, such as seeds and/or hard-cased fruits, as major dietary items.

**Kenyapithecus** The genus *Kenyapithecus* is retained for the rather later (14 Ma) species *K. wickeri*, known from part jaws and isolated teeth from Ft. Ternan, Kenya. The maxilla exhibits marked canine fossae, a relatively low and anteriorly positioned origin for the zygomatic process above M₁, little extension of the maxillary sinus into the inflated alveolar process (in contrast to *Equatorius*) and a relatively highly arched palate. The upper incisors are markedly heteromorphic, with I² much smaller than I₁, which is reinforced by strong lingual marginal ridges extending across the base of the crown surface. The upper canines exhibit marked dimorphism: robust, tall, and externally rotated in presumed males, more conical in females. P₄ is relatively broad with subequal cusps; M₁ is quadritubercular and lacks a lingual cingulum, while M₂ is similar but larger. The postcanine teeth are closely packed, with low cusp relief and appreciable wear.

The mandible, considered female, displays a shallow symphysis, sharply retreating at 30°–40° to the alveolar margin, with pronounced inferior transverse torus extending to below the mesial root of M₁, rather weak superior torus and long sublingual planum, a short incisor row, and a robust, comparatively shallow but thick corpus. As reconstructed, the mid-lower face overall was broad and flat,
with wide cheeks and a short snout (Andrews 1971; Walker and Andrews 1973). The relatively tall-crowned lower canine bears only a slight lingual cingulum, while P\textsubscript{3} is obliquely set and sectorial, with a distinct honing facet for the upper canine, and P\textsubscript{4} bears prominent mesial and smaller, lower distal cusps; the poorly known lower molars apparently lack buccal cingula.

The narrow anterior dental arcade, procumbent incisors with curved roots, anteriorly positioned zygomatic origin, restricted maxillary sinus and mandible with markedly sloping symphysis, pronounced inferior transverse torus and shallow, robust corpus differentiate \textit{K. wickeri} from most other African hominoids. Andrews (1971) and Walker and Andrews (1973) interpret these traits as a functional set adapted for powerful chewing activity with a strong lateral grinding component and pronounced incisal action. There are similarities, especially in the anterior dentition, with \textit{Afropithecus} and, as with that genus, pithecines have been proposed as the most plausible dietary analogues (Leakey and Walker 1997; McCrossin and Benefit 1997). This model of \textit{K. wickeri} as a sclerocarp feeder, exploiting hard-cased/hard-stoned fruits, seeds and nuts, is compatible with reconstruction of the Ft. Ternan environment as drier and more seasonal than many earlier African sites, predominantly closed-canopy woodland with both open country and forested conditions nearby (Andrews 1996; Andrews et al. 1997).

\textbf{Otavipithecus}  Broadly contemporary at $13 \pm 1$ Ma, the more southerly \textit{Otavipithecus namibiensis} is known from a part mandible and frontal bone from Berg Aukas, Namibia. The incisor region of the mandible is narrow and the symphysis reinforced by a short inferior transverse torus. The corpus is robust, of constant depth, and relatively long, with the ramus originating behind M\textsubscript{3} and with a distinct retromolar space. Premolar and molar cusps are inflated and bunodont, mesial, and distal foveae small, and enamel thin. The frontal bears superciliary ridges rather than a transverse torus, with marked temporal ridges adjacent to glabella, an extensive frontal sinus and relatively wide interorbital dimensions. The narrow incisor region does not support a niche of specialized frugivory, while the thin enamel and minimal wear differential on the molar teeth point to a nonabrasive diet; there are no obvious dental adaptations to folivory. Conroy argues that \textit{O. namibiensis} probably subsisted on a range of plant foods that required little preparation by the anterior teeth prior to chewing.

\textbf{Samburupithecus}  A later Miocene (9.5 Ma) large-bodied species from the Samburu Hills of north central Kenya named \textit{Samburupithecus kiptalami} is known only from one left maxilla with P\textsubscript{3}–M\textsubscript{3} crowns and the canine alveolus (Ishida and Pickford 1997). The palate displays a marked arch, a shallow
postcanine fossa, a low origin for the zygomatic root, and invasion of the zygomatic process by the extensive maxillary sinus. The nasal floor has a sharp margin, and the tooth row is straight from the canine alveolus to M\(^3\). The three-rooted premolars have elongated crowns with coequal main cusps, while the molars display inflated, bunodont cusps and thick enamel. \(S.\) kiptalami’s affinities are unknown; some workers consider it to show some similarities with the gorilla, although there are also undoubted differences, e.g., the size of lingual cingulae.

Other, later taxa, closer to the Mio-Pliocene boundary (\(Sahelanthropus, Orrorin\)) are as yet only incompletely described but are claimed as basal hominins (see Volume 3 Chapter 6 by Senut). The known time-span of \(Pan\) has recently been extended by the recovery of four fossil teeth (\(r\) and \(l\) \(I_1, l\) \(M_1, r\) \(M_3\)) from Middle Pleistocene deposits of the Kapthurin Formation of the Tugen Hills, Kenya, within the eastern Rift (McBrearty and Jablonski 2005). The broad, spatulate incisors bear deep mesial and distal foveae separated by a prominent lingual tubercle and the molars are low crowned, while all teeth exhibit thin enamel. The large hypocone on \(M^1\) suggests \(P.\) troglodytes rather than \(P.\) paniscus, although McBrearty and Jablonksi are cautious in attributing specific identity, preferring assignment to \(Pan\) sp. indet. The finds date from around 0.5+ Ma, and the site lies some 600 km east of the present chimpanzee range. The teeth were discovered close to localities yielding part mandibles of \(Homo\) (\(H.\) erectus or \(H.\) heidelbergensis/\(H.\) rhodesiensis) pointing to sympathy, and suggesting that adaptive scenarios which reconstruct differentiation of chimpanzee and hominin populations through the Rift Valley acting as an isolating barrier are unlikely to be correct. Fossil evidence for bonobos and gorillas is entirely lacking.

### 7.7.1.3 Europe and Asia Minor

While there is limited evidence dating from 15 to 17 Ma, the bulk of the European and west Asian hominoid material is from Later Miocene sites between 6 and 12 Ma. Many of these suggest subtropical seasonal forest or woodland as the dominant habitat; there is evidence of swamp conditions at some sites, while possibly harsher, more open environments are indicated at sites in Greece and Turkey yielding \(Ouranopithecus\) and \(Ankarapithecus\), respectively.

**Griphopithecus** The earliest evidence (13.5–17 Ma) consists of several low-crowned, large-cusped, and thick-enamed molars from Germany, Austria, and Slovakia assigned to \(Griphopithecus\), a genus better known from Turkey, where a mandible with cheek teeth from Candir and maxillary teeth from Pasalar dating
from c. 15 Ma are referred to *G. alpani* (Alpagut et al. 1990). Like their European counterparts, the lower molars have bunodont, thick-enamed cusp, and well-developed buccal cingula. The upper central incisors are relatively narrow with a distinct median lingual pillar, and the male canines, especially in the upper jaw, are robust and comparatively low crowned. The mandible is strongly constructed, with both a prominent superior and well-developed inferior transverse torus at the symphysis and a long, shallowly inclined planum alveolare.

Heizmann and Begun (2001) suggest that *Griphopithecus* evolved from an *Afropithecus/Heliopithecus*-like thick enamed ancestor and that this feature, together with the associated trait of low-dentine penetrance, was crucial to the expansion and success of dentally modern hominoids in the more seasonal Middle and Upper Miocene habitats of Eurasia. However, at least one successful European form—*Dryopithecus*—had comparatively thin enamel, pointing to this as either a secondarily derived trait evolved from a thicker-enamed European ancestor or that the genus represents a second hominoid radiation into Europe.

*Dryopithecus* The thin molar enamel of *Dryopithecus* results in frequent dentine exposure, especially on the cusps, which lie close to the crown margins around a broad, shallow fovea. Below the narrow, low P₃ crown enamel extends onto the anterior root, pointing to at least partial honing against the upper canine (Begun 2002), while the upper central incisors are narrow and high crowned. Larger (presumed male) mandibles are generally more robust than in the Early Miocene East African forms but are not as strongly built as more thickly enamed taxa (Begun 2002). All species in which evidence is available display a relatively high root for the maxillary zygomatic process. Compared with *Proconsul* and other Early Miocene forms, *Dryopithecus* has reduced cingula, relatively short lower molars (Szalay and Delson 1979), expanded occlusal surfaces, and reduced molar crown flare, and a mandibular symphysis reinforced by a prominent inferior transverse torus.

The type species, *D. fontani* (11–12 Ma), is among the less well-known cranially but appears to have been about chimpanzee sized. It has rather broader canines than other species, frequent cingula on the lower molars, while in larger mandibles the corpus shallows markedly from the symphysis toward the rear, unlike other species.

The later (9.5–10 Ma) *D. brancoi* is much better known (Figure 7.9). Comparable in tooth size to *D. fontani*, it differs in its labio-lingually thicker incisors, narrower canines, reduced molar cingula, and more tapered M₃. The mandibles have weak symphyseal tori but the corpus is reinforced below M₁–M₂ by a lateral eminence. Cranial morphology is comparatively well known from several incomplete specimens from Rudabanya, Hungary (Begun and Kordos
1997; Kordos and Begun 2001). The braincase is elongated, with a flat frontal displaying moderate postorbital constriction, and strong anterior temporal ridges; supraorbital reinforcement is weak, but there is a fair-sized frontal sinus. At the skull rear inion is relatively highly positioned, while the mandibular fossa is transversely deep, with marked entoglenoid and postglenoid processes. The face is moderately projecting and deflected downward. The maxillary sinuses are larger than those of the East African fossils and the nasal aperture has a broad base with subvertical sides. Again, in contrast to the early East African specimens where preserved, the subnasal floor is stepped, with the rear of the subnasal alveolar process extending over the palatal process of the maxilla and an incisive canal is present. The long and projecting premaxilla is sagittally and transversely convex.

*D. brancoi*’s smaller contemporary, *D. laietanus*, is known from several sites in NE Spain, including Can Llobateres, which has yielded a partial skeleton and
possibly associated cranium (CLI-18800) (Begun 2002). Mandibular teeth resemble those of *D. brancoi* but the premolars are relatively smaller, the molar cusps more rounded, and expanded around the occlusal margins, and M3 less tapered. CLI-18800 preserves the upper dentition: the incisors are like those of *D. brancoi*, the canines have strongly curved roots and narrow crowns, and the molar teeth increase in size posteriorly. In its known craniofacial structures, *D. laietanus* shows many similarities with *D. brancoi*, including periorbital and maxillary morphology, the supraorbital region, and frontal sinuses. However, there are also contrasts, with CLI-18800 displaying a very high root for the zygomatic process and a relatively deep and flatter anterior aspect of the zygoma. The upper incisor row is more strongly curved and the premaxilla is strongly biconvex. Overall the facial profile of *D. laietanus* is more concave.

The slightly earlier (c. 10.5 Ma) *D. crusafonti*, also from NE Spain, is known only from a mandible and some isolated teeth. Dentally slightly larger than *D. laietanus*, it is distinguished by comparatively broad upper canines, relatively longer upper premolars than in *D. brancoi*, and relatively broader upper molars than *D. laietanus*. The mandible combines an exceptionally robust corpus with comparatively small tooth crowns; the symphysis is reinforced by a strong inferior transverse torus and the corpus bulges laterally below M1–M3.

Overall, *Dryopithecus* dental features such as relative shearing crest development (Ungar 1996; Kay and Ungar 1997), pitting incidence of >35%–<40% (Ungar et al. 2004), thin enamel and dentine penetrance, and molar flare (Singleton 2003) all accord with reconstructions of the genus as a frugivore, probably primarily a soft-fruit feeder, in mildly seasonal subtropical forests (Andrews 1996). This is also consistent with mandibular corpus cross section which in many cases is like that of similar-sized *Proconsul*, although larger *Dryopithecus* specimens resemble *Sivapithecus* in their shallow, more robust and almost triangular section below M3—proportions particularly effective in resisting torsion (Brown 1997) and raising the possibility of a tougher, more fibrous component in the diet of larger individuals to sustain their greater body bulk.

**Pierolapithecus** The recently discovered Middle Miocene (12.5–13 Ma) partial skeleton of *Pierolapithecus catalaunicus* from Barranc de Can Vila 1, Hostalets de Pierola, Barcelona, Spain provides extensive new evidence of European hominoids (Moya-Sola` et al. 2004). The specimen (IPS 21350) includes a virtually complete face and lower frontal, the upper dentition and much postcranial material. The *Pierolapithecus* face is exceptionally prognathous compared with other Middle and Later Miocene Eurasian hominoids and, as such, is reminiscent of *Afropithecus*. It is low, with slender superciliary arches merging into a moderately swollen glabella region, below which the upper face is transversely flat.
The orbits are broader than high, and the interorbital distance wide. In profile, the nasal bones are concave and salient over the wide piriform aperture, the inferior margin of which is well anterior to the tips of the nasal bones. The nasoalveolar clivus is high, convex, and markedly projecting. The deep, laterally expanded and strongly constructed zygomatics slope antero-inferiorly, with their root originating high above the alveolar margin of M1. The palate is short, wide, and deep, with the anterior dentition arcuate. I1 is low crowned and procumbent, the large C low crowned and compressed. P3 and P4 are of similar size with reduced cusp heteromorphy, while M1 and M2 crowns are long and relatively narrow, with M3 reduced. The lingual cusps are situated toward the crown edge and all the cheek teeth lack cingula.

This combination of dental and facial features distinguishes *Pierolapithecus* from all other Miocene hominoids; contrasts with *Dryopithecus*, include the lower, more prognathous face, more anteriorly positioned zygomatic roots, shorter wider palate, larger, low-crowned anterior teeth, and relative molar crown size. In the view of Moyà-Sala et al. (2004) shared derived features of the two genera, which link them to extant great apes, include upper facial flatness, nasal bone projection and aperture form, a high zygomatic root, high nasoalveolar clivus, and deep palate. The postcranium also reveals shared derived traits with extant hominoids: a broad, shallow thorax and stiff lumbar region to the trunk, a dorsally positioned scapula, and apelike carpal bones with ulnar retreat from the wrist joint and evidence for the fibrous capsule of a semilunar meniscus, providing enhanced wrist abduction and supination. However, these are combined with metacarpal and phalangeal features, indicating the hand was palmigrade during locomotion and the fingers short, as in monkeys, pointing to a dissociation in ape phylogeny between orthograde posture and climbing, and below-branch suspension. The latter appears to have evolved later than the former and may well have arisen independently in several distinct lineages.

Similarities in craniofacial form and dental proportions with *Afropithecus* suggest a sclerocarp feeder, consistent with reconstructions of the *Pierolapithecus* habitat, based on the associated fauna, as wooded and relatively humid. Several Later Miocene hominoids (*Ouranopithecus*, *Ankarapithecus*, *Graecopithecus*) exhibit notably more ruggedly constructed crania and derive from drier, more fluctuating, and possibly more open environments.

**Ouranopithecus** *Ouranopithecus macedoniensis*, represented by several jaws, teeth, and a fairly complete face from the c. 9-Myr old northern Greek sites of Ravin de la Pluie, Xirochori, and Nikiti 1 (*Figure 7.10*), is the largest European hominoid known, with males estimated to have been about female gorilla size. The face is strongly reinforced in its mid and upper portions, pointing to
powerful masticatory forces: the rectangular orbits are separated by a broad, stout interorbital area and laterally bounded by massive orbital pillars, while the nasal aperture is flanked by thick naso-maxillary (canine) pillars with broad, deep canine fossae beyond. The zygomatic region is very deep with, unusually for hominoids, a low origin for the zygomatic root on the maxilla. The premaxilla is strongly built, comparatively long, and markedly convex sagittally and transversely, with clivus, subnasal fossa, and incisive canal resembling *Dryopithecus*. The supraorbital torus bulges moderately above the orbits but less so medially; the strongly marked temporal lines indicate powerful anterior *temporalis* fibers.

The palate is deep and anteriorly broad; the incisors are set in an arc, offset from the other teeth by a pronounced diastema. The upper incisors are markedly heterodont, with the central teeth thick and broad, the laterals narrow and peg
shaped; lower incisors are tall and slightly flared. Larger (male) canines are tall and laterally compressed but small when compared with molar size, while smaller (female) canines are very low crowned and almost premolariform.

Brown (1997) considers such low canines compatible with a greater degree of lateral mandibular movement than in modern apes. Both upper and (especially) lower anterior premolars are triangular, the latter with a prominent mesial beak and lacking an anterior vertical honing facet from the upper canine. Instead wear occurs on the tip and along the protocristid, a feature judged reminiscent of *A. afarensis* (de Bonis and Koufos 2001). P^4^ resembles P^3^ in shape—another claimed *Australopithecus*-like trait and unlike other fossil and extant apes—while P^4^ is molarized, being elongate and with a high talonid. The molars are relatively large, with M^2^ and M^3^ the largest teeth; individual cusps are expanded, and both upper and lower rear molars possess accessory cusps. There is a marked wear gradient along the molar row: enamel is very thick and the cusps are worn almost flat before dentine appears.

Mandibles are robust, with thicker corpora than in extant apes, and larger (male) specimens are deep as well as thick. The symphysis is strongly reinforced by a broad, long planum and prominent inferior transverse torus. The lateral eminence originates opposite M^3^ and the ramus ascends between M^2^ and M^3^; The gonial region is extensive and displays strongly marked ridging for the superficial masseter and medial pterygoid muscles. The condyle is relatively large and sagittally strongly convex.

Facial structure, muscle impressions, dental structure and proportions, occlusal morphology, and attritional wear gradients in *Ouranopithecus* all contrast with *Dryopithecus*, and suggest extremely powerful masticatory activity, especially of the cheek teeth. This is further supported by microwear patterns which reveal a very high ratio of pits compared with striations on Phase 2 facets of the cheek teeth (at >58% the highest of any fossil ape studied) and distinct wear on the incisors (Ungar 1996; Ungar et al. 2004), while the *Ouranopithecus* Shearing Quotient (and so relative shearing-crest development) is lower than that of any modern hominoid or other European fossil ape studied (Ungar and Kay 1995; Kay and Ungar 1997). In all these features, *Ouranopithecus* resembles extant hard-object feeders, pointing to exploitation of a similar niche—perhaps seeds, nuts, roots and tubers, and other terrestrial vegetation. There are obvious resemblances here to some reconstructions of early (Plio-Pleistocene) hominine dietary niche(s), although these are unlikely to mirror any specially close phylogenetic link. De Bonis and Koufos (1997) argue that this model of *Ouranopithecus’* diet is consistent with reconstruction of its open habitat (de Bonis et al. 1992), although Andrews (1996) and Andrews et al. (1997) urge caution,
considering the overall fauna to be undiagnostic other than indicating a strongly seasonal, possibly harsh, environment.

The poorly known *Graecopithecus freybergi* (von Koenigswald 1972), based on a single mandible from Tour la Reine, Pygros, Greece and dated around 6.5—8 Ma, is often regarded as congeneric or even conspecific with *O. macedoniensis* (Martin and Andrews 1984; Andrews 1996), although Begun (2002) makes a strong case based on molar size and mandible proportions, for its retention as a separate taxon.

**Ankarapithecus**  *Ankarapithecus meteai* (Ozansoy 1965), from sites in the Sinap Formation of Anatolia, Turkey dated at c. 10 Ma, is a strongly built form known from cranial material including the type mandible and a partial face (Figure 7.10), together with undescribed postcrania. The face is tall, and markedly prognathic in both mid-face (unlike *Sivapithecus* and *Pongo*) and premaxillary regions. The clivus is biconvex, with large, low-crowned and labio-lingually thick central incisors and smaller lateral incisors. Male upper canines are relatively low crowned, and their roots form strong juga converging on the broad nasal aperture, with relatively shallow canine fossae beyond. The palate is deep, with the root of the zygomatic process set comparatively high above M1 and into which the large maxillary sinuses extend. The vertically orientated, laterally flaring zygomatic process imparts strong anteriorly and laterally directed components to masseteric action, while the deep temporal fossa allows for a powerful temporalsis muscle. The subnasal fossa is stepped and the incisive fossa large. The orbits are square, with a narrow interorbital space, very long nasal bones, broad, rounded orbital pillars, and prominent anterior temporal lines. Rather surprisingly in view of the rugged mid- and lower face, the superciliary arches above the orbits do not form a true torus.

The massive mandible is strongly buttressed, with a very deep, narrow, and vertical symphysis and the inferior traverse torus extending to the level of M1. The rear of the corpus is very thick and the ramus broad. The lower incisors are labio-lingually thick, narrow, and tall crowned, set almost vertically in a straight line between the low-crowned canines, which in males are more massive than in *Dryopithecus* or *Ouranopithecus* in basal section, while female canines are more premolariform, as in the latter genus. P3 is large, oval and elongated, with a large mesial beak comparable to *Dryopithecus* but smaller than *Ouranopithecus*, P4 is large and relatively broad, while M1 is small relative to M2. Upper and lower molars are broader relative to their lengths than in other later Miocene forms; their occlusal surfaces have broad, flat cusps, shallow basins and lack cingula.

Andrews and Alpagut (2001) provide a valuable functional analysis of this taxon as a hard object feeder in dry seasonal subtropical forest (Andrews 1996),
illuminating aspects of *A. meteai* morphology and also that of other Miocene hominoids. In many features *Ankarapithecus* resembles *Ouranopithecus* dentally in the large, low-crowned and worn incisors, large cheek teeth, thick enamel, flat occlusal wear, poor shearing crest development; and also in aspects of facial architecture and mandibular reinforcement. However, there are also differences: the supraorbital region differs, the *Ankarapithecus* interorbital region is narrower as are the lateral orbital margins, the mid-face is more prognathic, and the zygomatic root originates higher on the maxilla, while the mandibular corpus is massively thickened under the rear molars, so that it is actually broader than deep, whereas that of *Ouranopithecus* is deeper and narrower in cross section.

**Oreopithecus**  *Oreopithecus bambolii*, a comparatively large-bodied form dating from 6 to 7 Ma, is represented by multiple specimens from northern Italy, including a largely complete skeleton, making it the best-known European fossil primate. This has not prevented protracted debate about its affinities, although in recent years there has been a growing consensus that *Oreopithecus* is a primitive hominoid. Whatever its phyletic status, it is clear that *Oreopithecus* differs markedly in adaptive features and inferred niche from other Late Miocene European and West Asian hominoids.

The *Oreopithecus* skull combines a relatively small, low, but globular neurocranium with a deep, broad, and moderately projecting face that reflects anterior placement and projection of the mid-face and nasal region, for the premaxilla and clivus are short and comparatively vertical. The supraorbital torus is well developed, the interorbital region broad, and the nasal bones short and salient. Strong canine pillars are bounded by shallow canine fossae, while the zygomatic process is comparatively deep with a low, anteriorly placed root originating above P4/M1, and the zygomatic arch is long, flaring, and upwardly curved posteriorly. The alveolar region has deep but restricted sinuses. The saddle-shaped articular eminence is broad and long, with a large entoglenoid process. The articular and tympanic portions of the temporal are not fused, but the temporal petrous is hominid-like in its shallow, indistinct subarcuate fossa. The mastoid is broad and continuous with the extensive, strongly marked nuchal area. Zygomatic flare and mandibular proportions point to powerful, fleshy temporal muscles, as do the deep nuchal and sagittal crests, meeting high on the skull rear; the sagittal crest continues well forward before dividing into two prominent anterior temporal lines.

The strongly built mandible is large, with corpus height decreasing slightly along the cheek teeth row, and with strong reinforcement provided by the pronounced lateral eminence below the molar region. The ramus is broad and high with an expanded gonial region. There are pronounced markings for the
masseters and medial pterygoids, while the condylar processes, below the broad and convex condyles, display strong markings for the lateral pterygoids.

Many of these cranial features can be considered representative of the primitive catarrhine morphotype (Harrison 1986); the exceptions are those features of the maxilla, zygomatic region, and mandible summarized above that can be related to masticatory power (see below). The dentition is highly derived: the incisor teeth are small overall and vertically implanted; those in the mandible are labio-lingually compressed, while the uppers are heteromorphic with I\(^1\) exhibiting a distinctive projecting lingual cusp. Canines are basally stout but not very tall and only loosely interlock, with a diastema small or absent. Larger (male) upper canines are strongly compressed and with a sharp rear edge, the lowers are more rounded in section; smaller (female) upper canines are rather incisiform. Much canine wear is from the tips; there was some C\(^1\)/C\(_1\) honing, and some larger upper canines show evidence of slight honing against P\(_3\), but smaller ones lack this, and the anterior face of C\(_1\) did not hone against I\(^2\). The lower premolars are bicuspid, with P\(_3\) oval in outline and P\(_4\) more rectangular; upper premolars are oval with subequal cusps. Upper and lower molars are elongated, and bear tall, spiky cusps with deep notches between. Besides the four main upper molar cusps, a metaconule is positioned centrally on the crista obliqua, often linked to the hypocone by a crest. On the lower molars, the protoconid and hypoconid (buccally) and meta- and entoconid (lingually) are joined by sharp crests to a well-developed centroconid on the cristid obliqua, so mirroring the upper molars in their distinctive occlusal pattern. The hypoconulid is frequently split into several smaller cusps. M\(_1\) usually bears a small paraconid, which is rarely present on M\(_2\) and never on M\(_3\). M\(_1\) and M\(_2\) are subequal in size, and M\(_3\) is the largest tooth.

Although it displays some primitive features, the postcranial skeleton is apelike, indicating a degree of orthogradly, forelimb suspension, and strong grasping capabilities in the feet. The thorax is broad, the lumbar region short, the iliac blades are short and broad, with a prominent anterior inferior iliac spine. The forelimbs are much longer than the hindlimbs and display multiple adaptations to stability and hyperextension at the elbow joint, and rotation in the forearm. There was a wide range of movement at the wrist, with short palms and long, curved fingers. The femora show weight-bearing adaptations at hip and knee, with flexible, wide-ranging movements at the ankle, and a short midfoot with long, strongly muscled digits including a powerful, opposed hallux.

Oreopithecus’ reduced anterior dentition, expanded cheek teeth with complex occlusal morphology, relatively deep and orthognathic face, small neurocranium combined with large ectocranial crests and powerful chewing muscles, and robust mandible all point to a specialized foliverous diet of bulky, relatively low
grade food items which, judging by postcranial morphology and proportions, it exploited largely via an underbranch milieu. This conclusion is reinforced by study of relative shearing crest development (Ungar and Kay 1995; Kay and Ungar 1997), which shows it to have the highest-shearing quotient of any catarrhine studied, substantially in excess of any extant or other fossil hominoid. Further support is provided by dental microwear patterns, with a very low proportion (17%) of pitting on Phase II facets (Ungar 1996; King 2001; Ungar et al. 2004), consistent with extreme folivory. This wholly accords with reconstructions of Oreopithecus’ paleoenvironment, which indicates lowland mixed broad-leaved and coniferous forest, with bushes, ferns, and sedges accumulated under swampy conditions (Andrews et al. 1997; Harrison and Rook 1997). However, see Alba et al. (2001) for an alternative—and to our mindless convincing—interpretation of the Oreopithecus skull based on biomechanical constraints associated with orthograde posture and bipedalism. These authors derive Oreopithecus cranial morphology from a Dryopithecus-like ancestor by a process of neoteny.

7.7.1.4 South and East Asia

Sivapithecus The best-known Asian fossil ape genus is Sivapithecus, from Late Miocene (8.5–12.7 Ma) deposits in the Siwalik Hills of India and Pakistan and including material assigned to Ramapithecus prior to the 1980s. Sivapithecus is rare throughout the Siwalik record, comprising only c. 1% of the mammalian community (Ward 1997). Despite this, aspects of its cranial morphology are relatively well known through discoveries over the last three decades; in particular, a partial skull (GSP 15000) from Potwar, Pakistan, provides much information on facial and gnathic morphology (Figure 7.11). It consists of the left side of the face with zygomatic arch, palate, mandible, and complete adult dentition (Pilbeam 1982).

The specimen indicates many similarities with Pongo in its overall facial proportions (but see below): the orbits are taller than broad and high-set, ovoid in outline, and the zygomatic foramina are large. The interorbital distance is very narrow, while the lateral orbital pillars are slender, especially sagittally, and there are distinct supraorbital ridges but no continuous torus.

Postorbital constriction is marked, with the anterior temporal lines strongly convergent, implying a well-developed sagittal crest in larger individuals. The frontal rises more steeply above the orbits than in extant nonhuman African apes and is orangutan-like in its contour and the absence of a frontoethmoid sinus.

The nasoalveolar clivus is long and strongly curved, intersecting the alveolar plane at a shallow angle and, as in the orangutan but in contrast to Dryopithecus,
Ouranopithecus, and Ankarapithecus, the nasal floor is smooth, with the premaxilla curving into the nasal cavity and joining the palatal process without a step; the incisive fossa and incisive foramen are both tiny, linked by a very narrow incisive canal. The long, medially convergent canine roots are externally rotated, while the deep zygomatic process is thin and *Pongo*-like in its flare. These features result in exceptionally prominent canine pillars reinforcing the robust anterior mid-face, and well-marked canine fossae lateral to them (Ward and Pilbeam 1983). The GSP 15000 mandible is deep and strongly built; the symphysis exhibits pronounced buttressing, the corpus is of fairly constant depth along the tooth row, while the ramus is high and broadest at the level of the occlusal plane, tapering slightly superiorly. The facial contour is *Pongo*-like in its marked concavity, nasoalveolar clivus projection, superoposterior slope to the zygomatic process and lateral orbital pillar, and the upward inclination of the zygomatic arch itself.

 Nonetheless, there are contrasts with the orangutan: the *Sivapithecus* mid-face is much longer, the nasal bones especially so, and the maxillary sinus is more restricted (Ward 1997). The mandible in particular contrasts with *Pongo* in most features other than its high ramus: while highly variable (Brown 1997), all *Sivapithecus* specimens display markings for the anterior digastric muscles which are absent in orangutans (see above), and the inferior transverse torus does not extend as posteriorly as in *Pongo* (Brown 1997). There are also contrasts in corpus cross section: *Sivapithecus* specimens have robust, broad corpora, ovoid below the premolars, more triangular below M₃, with smaller specimens relatively shallow, larger ones deeper. They show marked relief, with an intertoral sulcus.
near the lingual base below the cheek teeth, and a pronounced lateral eminence which continues to the base, whereas in orangutans the lateral eminence is usually a much less prominent swelling restricted to the upper part of the corpus. *Sivapithecus* mandible proportions have been interpreted as resisting sagittal bending loads on the balancing side and pronounced torsional and shearing loads on the working side associated with powerful molar action and, as a possible secondary factor, incisal biting (Kelley and Pilbeam 1986; Brown 1997).

Dental features accord with this interpretation: the upper incisors are strongly heteromorphous: the central teeth are very wide and spatulate with a heavily crenulated, extended lingual tubercle; the laterals are much narrower. The canines are moderately tall, compressed, and outwardly rotated, and display only limited dimorphism. P³ is larger than P⁴, while the mandibular premolars are broad, with P₃ expanded mesiobuccally and displaying only limited evidence of upper canine honing. The thickly enameled molars lack cingula and display expanded, bunodont cusps and so limited occlusal foveae. In the upper jaw, M² is the largest tooth, while M₃ is the largest of the relatively short and broad lower molars. The cheek teeth are closely packed, with clear interproximal wear facets and, despite the thick enamel, often display a pronounced wear gradient with, in older individuals such as GSP 15000, destruction of crown relief and extensive dentine exposure.

Overall, evidence points to *Sivapithecus* as a frugivore/hard object feeder, possibly nuts, seeds, bark, or hard-pitted fruits, requiring powerful mastication by the postcanine teeth. Earlier scenarios of the genus as an open habitat form, exploiting terrestrial vegetation, have been replaced by reconstructions of its environment as predominantly seasonal tropical or subtropical closed canopy forest or woodland, albeit with patchiness and expanding areas of more open grassland in the later phase of its presence in the Siwalik record (Andrews et al. 1997; Ward 1997). This shift probably reflects broader climatic changes that, through the contraction and break up of its forested habitat, eventually resulted in the extinction of *Sivapithecus*.

Three *Sivapithecus* species are recognized, differentiated primarily on dental proportions: *S. sivalensis* from Siwalik sites dating between 8.5 and 9.5 Ma is the type species. *S. indicus*, represented by GSP 15000 and other material, is earlier (10.5–12.5 Ma), and with absolutely rather smaller teeth than *S. sivalensis*, but with a proportionately larger M₃ compared with M₂, and with a rather shorter premaxillary region. A humerus with a retroflexed and mediolaterally strongly curved shaft and a prominent deltopectoral crest is assigned to this species. *S. parvada* is a recently recognized, appreciably larger form, dating around 10 Ma. Its I¹ is particularly wide relative to its breadth, the premolars, especially the lower ones, are expanded relative to molar size, and M₃ is again much larger.
than M₂. The mandible’s symphysis and anterior corpus region are exceptionally deep while a humerus referred to *S. parvada* broadly resembles that of *S. indicus* but is much bigger, implying larger body size overall.

**Gigantopithecus** Many workers regard the Asian genus *Gigantopithecus*—known from massive mandibles and individual teeth from the Late Miocene/Pleistocene of southern China, Vietnam, and the Siwaliks of India and Pakistan—as closely related to *Sivapithecus*. While extremely large, it is characterized by a reduced anterior dentition, with relatively small lower incisors and low-crowned but basally large canines without honing facets, strongly worn down from the tip and functionally incorporated in the premolar/molar rows. The expanded premolars are strongly molarized: P³ is bicuspid with a large talonid and is larger than P⁴, which is almost square with a large trigonid taller than the talonid. The upper molars are almost square, the lowers elongated; all have very thick enamel, high crowns, and low cusps. The symphysis is reinforced by a moderate superior transverse torus and a much more extensive inferior torus that may extend as far back as M₁.

*Gigantopithecus giganteus* is known from specimens found from the 1960s onward from Haritalyangar and other Siwalik sites, especially a mandible CYP359/68 (Pilbeam et al. 1977), usually considered late in the Siwalik sequence at <7 Ma (Johnson et al. 1983). However, Pillans et al. (2005) argue for a date of 8.6 Myr for the mandible, and report a newly discovered M² from the earlier (8.85 Ma) HD site at Haritalyangar which also yielded an incisor assigned to *Sivapithecus*, demonstrating sympatry of the two genera. Fossils of the more recent *G. blacki* are even larger than those of *G. giganteus*, with more pronounced molarization of the premolars, so that P₃ is more distinctly bicuspid and P₄ longer, while the molars are higher crowned, with low, expanded cusps, and multiple accessory cusplets. The mandibular symphysis is long and powerfully buttressed; the corpus is strongly reinforced by a thick lateral torus originating below M₁ and corpus depth increases posteriorly (Weidenreich 1945). Given these and other contrasts between the South and East Asian species, Cameron has recently reassigned the Haritalyangar mandible to *Indopithecus* (Cameron 2001, 2003; Pillans et al. 2005).

The size of *Gigantopithecus* fossils almost certainly precludes arboreality, and most reconstructions are of a ground-dwelling pongin exploiting a low grade, bulky diet—perhaps bamboo (present at both Siwalik and Chinese localities), although more varied diets, including fruits, have been suggested (Ciochon et al. 1990). Whatever the details, dental and gnathic features clearly indicate an extremely powerful masticatory apparatus with a premium
on occlusal crushing and grinding by the cheek teeth, together with an anterior dentition capable of generating considerable bite forces for cropping food items. The short, premolariform canines, together with the mandible’s symphyseal and corporal proportions, point to the generation of powerful torsional and shearing forces during food processing. The same environmental changes that reduced forest and woodland cover, leading to the extinction of *Sivapithecus*, may well have favored the evolution of *Gigantopithecus* as a form better adapted to more open conditions and the exploitation of terrestrial vegetation (Ward 1997; Pillans et al. 2005).

**Lufengpithecus**  *Lufengpithecus* is another large-bodied Asian pongin, represented by extensive material, including several distorted crania and numerous teeth, from the Late Miocene (c. 8 Ma) of Lufeng, Yunnan province, southern China. Dental metrics indicate a degree of sexual dimorphism rather greater than that of any extant ape (Kelley and Qinghua 1991; Kelley 1993). Cranial morphology contrasts with *Sivapithecus* in the relatively shorter mid-face, square orbits, broad interorbital and glabellar regions, and more prominent supraorbital torus above the medial superior orbital margins, resembling more the *Dryopithecus* pattern. The nasoalveolar clivus is relatively short, and the mandibular symphysis strengthened by a moderate superior and strongly developed inferior tranverse torus. The corpus is narrow and columnar with little sculpting or buttressing (Brown 1997), presumably resistant to vertical occlusal forces, but less well adapted to torsional or shearing forces generated by the cheek teeth.

The upper central incisors are tall and narrow but relatively thick and buttressed by a median pillar on the lingual surface, while the lower incisors are relatively narrow and moderately procumbent. Presumed male canines, especially the lowers, are tall, sharply tapering and relatively slender. Molar enamel is moderately thick, and the crowns are narrow with the cusps situated toward the rim so that the foveae are extensive, and the occlusal surface bears a complex pattern of crenulations. Overall molar occlusal morphology is remarkably like that of the orangutan (Ward 1997), suggesting a frugivorous niche, as do the similarities with *Dryopithecus*. Paleoenvironmental indicators point to swampy moist tropical forest conditions, with ferns and epiphytes in which—again unlike *Sivapithecus*—*Lufengpithecus* was common, representing >33% of the mammalian fossils (Andrews et al. 1997).

Kelley (2002) recognizes three species of *Lufengpithecus*: *L. lufengensis*, the type species, is the best known; *L. keiyuanensis* and *L. hudienensis* have smaller postcanine teeth and rather greater molar cingulum development than *L. lufengensis*, with *L. keiyuanensis* possibly also having thinner enamel.
**Khoratpithecus**  The recent discovery of teeth and a part mandible from Middle and Upper Miocene deposits at sites in northern Thailand, and so within the geographical range of Pleistocene *Pongo*, sheds further light on orangutan ancestry (Chaimanee et al. 2003, 2004). The finds have been assigned to the new genus *Khoratpithecus*, and display numerous dental and gnathic similarities with *Pongo*, as well as with *Lufengpithecus* and, to a lesser extent, *Sivapithecus*. *K. piriyi*(9–7 Ma) is a large form (estimated 70 to 80-kg body weight) from a locality in Khorat, NE Thailand, and represented by a mandible body with the left canine—right I₂ roots, and with the right canine and all cheek teeth crowns preserved on both sides (Chaimanee et al. 2004). The symphysis is strongly sloping, thicker in overall cross section than usual in *Pongo*, with a weaker superior traverse torus, shallow genial fossa, and strongly developed inferior torus that, while wide and extending to below the anterior part of the M₁ crown, is less posteriorly extensive than in the orangutan. While the geniohyoid muscle facets are distinct, *Khoratpithecus*, like *Pongo*, lacks any impression for the anterior digastric muscle. The corpus is uniformly deep, with a marked depression on the lateral surface below the C/P₃ region and thickening posteriorly, accentuated by a pronounced lateral eminence below M₃.

Judging by anterior jaw proportions and alveoli, the procumbent incisors were larger than in *Lufengpithecus* but smaller than *Pongo*, and arranged in a slightly convex arc. Enamel wrinkling, while present, is less complex than in the orangutan, the P₄ is shorter, and the molar cusps are more centrally located than in the modern ape. The site indicates a riverine setting with palms and Dipterocarps, together with proboscids, anthracotheres, pigs, rhinos, bovids, and rare hipparion, corresponding to the Upper Nagri/Lower Dhok Pathan Formation faunas in the Siwalik sequence.

An earlier species, *Khoratpithecus chiangmuanensis* (13.5—10 Ma) is based on upper and lower teeth of a single individual from Ban Sa in the Chiang Muan basin (Chaimanee et al. 2003). Enamel wrinkling, markedly heterodont upper incisors, P₃ crown form, lack of molar cingula and comparable degrees of relative enamel thickness, and dentine penetrance again align it with *Pongo*, but it differs from the latter in its smaller central incisors, in the weaker median lingual pillar of the upper and lower incisors, in the greater buccal flare of the lower molar crowns, and in its less intensive enamel wrinkling.

Contextual evidence associated with *K. chiangmuanensis* indicates a mosaic of tropical freshwater swamps and lowland forest that contrasts with the temperate flora from Lufeng, instead resembling modern African habitats such as those in the southern Sudan around the source of the White Nile. Chaimanee et al. take this to indicate a Middle Miocene floral and faunal dispersal corridor linking South East Asia and Africa that may have been critical in hominoid dispersion.
Overall, *Khoratpithecus* closely resembles *Lufengpithecus* in its dentition, but its similarities with *Pongo*, especially the absence of digastric fossae, point to closer affinity with the modern genus than any other fossil ape.

### 7.8 Summary and conclusions

Early Miocene hominoids already show considerable craniodental diversity, probably associated with dietary niche differentiation. Better known genera (*Afropithecus*, *Proconsul*) have contrasting morphologies but span a range of frugivory: *Proconsul* sp. and *Dendropithecus* were probably soft-fruit feeders, *Afropithecus* exploited hard-cased fruits. Among the less well-known genera, some (e.g., *Morotopithecus*, *Turkanapithecus*) bear cranial and/or dental similarities to *Afropithecus*, suggesting a hard-fruit niche, and that the most familiar cranial morphology—that of *Proconsul*—is not necessarily characteristic of many Early Miocene hominoids. *Limnopithecus* and *Micropithecus* appear to have been folivores on the basis of their dentition.

Middle and Later Miocene forms from the region (*Nacholapithecus*, *Equatorius*, *Kenyapithecus*, *Samburupithecus*) are thick enameled and probably hard-cased fruits and seed feeders, while environmental evidence suggests more open, rather drier, and more seasonal habitats. The more southerly *Otavipithecus* is an exception, with thin enamel and minimal dental wear, pointing to a soft-fruit diet.

By the Middle Miocene, hominoids are known from Europe and western Asia. The earliest (*Griphopithecus*) are thick enameled, as are many later genera which are also generally more robust cranially than *Proconsul*, some especially so. However, at least one successful genus (*Dryopithecus*) has thin enamel and an only moderately strongly constructed cranium, although occlusal area, inferior symphysial reinforcement, and, in larger individuals, mandibular cross section are expanded compared with *Proconsul*. *Dryopithecus* also contrasts with early East African fossils in its stepped nasal floor with the alveolar process overriding the palate and an incisive canal present. Overall, evidence suggests *Dryopithecus* primarily exploited soft fruits. The recently described, markedly prognathous *Pierolapithecus* is much more reminiscent of *Afropithecus* in its morphology and, as such, probably a sclerocarp feeder.

Broadly contemporary with the younger *Dryopithecus* species at 9–10 Ma, *Ouranopithecus* and *Ankarapithecus* are more strongly built forms whose cranial reinforcement, muscle markings, gnathic proportions, and dental features all point to impressive masticatory power and hard-object feeding, characteristics shared with the less well-known and rather younger *Graecopithecus*. Of about
the same age (7–8 Ma) is the (masticatory power apart) generally contrasting *Oreopithecus*. This genus retains many primitive cranial traits together with features making for enhanced chewing capability and a distinctive dentition adapted to specialized folivory.

A similar trend to more robust morphologies is seen in South Asian hominoids, although details differ. The early Upper Miocene *Sivapithecus* (8.5–<13 Ma) is broadly *Pongo*-like in many aspects of cranial morphology, including periorbital proportions, mid/lower facial prognathism and cheek orientation, nasal floor structure and the presence of a narrow incisive canal, and remains among the most convincing instances of a comparatively close phylogenetic link between fossil and extant hominoid taxa, although *Sivapithecus* is unlikely to be directly ancestral to the orangutan. Cranial and dental features of *Sivapithecus* point to a frugivorous/hard-object feeding niche in increasingly patchy, fragmented habitats. These same environmental shifts probably underlie the appearance of *Gigantopithecus* (*Indopithecus*) *giganteus* in the Siwalik record, the oldest example of which is sympatric with *Sivapithecus*. *Gigantopithecus* mandibles are massive and powerfully reinforced, while occlusal area is expanded through increased molar size, premolar molarization, and incorporation of the low-crowned, worn flat canines into the cheek teeth rows. These features are evident in *G. giganteus*, but even more pronounced in the later, East Asian, *G. blacki* (2–<1 Ma). Given its size, *Gigantopithecus* probably exploited bulky, low-grade terrestrial vegetation, with bamboo as the most likely predominant food source.

Also from East Asia and roughly contemporary with latest *Sivapithecus* and earliest *Gigantopithecus* is *Lufengpithecus*. In some respects, this is cranially rather reminiscent of *Dryopithecus*, contrasting with *Sivapithecus* in its broad upper face, supraorbital development and orbital proportions, thinner enamel, in its mandibular section which is less resistant to torsion and shear forces, and in its molar occlusal pattern which is more *Pongo*-like than that of *Sivapithecus*, all of which suggest a frugivorous diet.

The immediate ancestry of the extant great apes is obscure, while the entire phylogeny of hylobatids is unknown. Both the distribution and diversity of contemporary *Hylobates* and *Pongo* populations indicate that climatic changes and associated sea-level fluctuations have been major determinants of their evolution over the last 2+ Myr. *Pongo* in particular is cranially highly variable: there are appreciable differences between Bornean groups separated by major river barriers, as well as marked contrasts between these and Sumatran orangutans, leading to recent proposals for species-level distinction between the two island populations. All orangutans lack the anterior digastric muscle, thereby contrasting with virtually all other hominoids including *Sivapithecus*, which does, however, share with *Pongo* a distinctively aiorrhynchous cranial form.
The recently discovered Mid/Late Miocene *Khoratpithecus* from Thailand displays jaw and dental affinities with *Sivapithecus* and especially *Lufengpithecus* and *Pongo*, sharing with the last a lack of any indication of the anterior digastric muscle.

The notably klinorhyncous African apes contrast in this respect and exhibit multiple similarities that indicate a common cranial pattern differentiated by varying growth periods. Since a degree of aiorrhynchy seems common among fossil hominoids and differentiates hominoids from non-hominoids, the African ape condition appears derived. Despite the comparatively full Miocene fossil record, there are no especially convincing candidates for modern African ape ancestry, although *Samburupithecus* is a potential link. The sequence and timing of splitting of the gorilla, chimpanzee, and hominin clades is uncertain, although some evidence (dental, gnathic, temporal) suggests that *Gorilla* is more primitive and *Pan* more derived. Differentiation of bonobos and chimpanzees, and of east and west gorilla populations perhaps occurred in the Late Pliocene/Pleistocene. Recently recovered fossil teeth provide evidence of *Pan* (probably *P. troglodytes*) from c. 0.5 Ma in the East African Rift.

While there are broad associations between African ape diet and cranial form, more detailed analyses fail to show an exact correspondence, in part because of dietary variability and also because of cranial variation. Evidence suggests that cranial features are less closely determined by diet than are characteristics of the dentition.

Bonobos show reduced levels of sexual dimorphism in craniodental features compared with chimpanzees, perhaps reflecting differences in sexual behavior and social organization. Surprisingly, bonobos are *relatively* the most robust of the African apes in some cranial traits, gorillas *relatively* the most gracile. When viewed in the broader context of the hominoid fossil record, all the living great ape crania are comparatively lightly constructed, raising issues about their representative nature or otherwise for the functional and adaptive modeling of earlier hominoids.

**Acknowledgments**

We thank Ms. J. Manghan for photographing specimens, and the Curators of the Grant Museum of Zoology, University College London, and the National Museums of Kenya for access to specimens in their care. Special thanks are due to Malgosia Nowak-Kemp of the Oxford University Museum of Natural History for facilitating access to specimens there. We are grateful to the following colleagues who kindly provided photographs: P. Andrews, D. Begun, J. Kappelman,
T. Koppe, S. Moyà-Solà, D. Pilbeam, E. Sarmiento, and especially C. Groves and H. Turni who generously provided multiple images; that we were unable to use all of these as illustrations here does not lessen our debt to them for allowing us to access their extensive documentation of hominoid cranial morphology.

References


Dean MC (1985) Comparative myology of the hominoid cranial base. II: The muscles of the prevertebral and upper pharyngeal region. Folia Primatol 44: 40–51


Delattre A, Fenart R (1960) L’Hominisation du Crane. CNRS, Paris


Johnson GD, Opdyke ND, Tandon SK, Nanda AC (1983) The magnetic polarity stratigraphy of the Siwalik Group at Haritalyangar (India) and a new last appearance datum for Ramapithecus and Sivapithecus. Palaeogeog Palaeoclim Palaeoecol 44: 223–249
Kelley J (1993) Taxonomic implications of sexual dimorphism in Lufengpithecus. In:
postcranium: A developmental perspective. Proc Natl Acad Sci USA 96: 13247–13252


Schwartz JH (1997) Lufengpithecus and hominoid phylogeny. Problems in delineating and evaluating phylogenetically relevant


Weidenreich F (1945) Giant early man from Java and South China. Anthropol Pap Am Mus Nat Hist 40: 1–134


Wood Jones F (1929) Man’s place among the mammals. Arnold, London
Abstract

Improvements in the primate fossil record, and in methods of data acquisition and analysis, have set the stage for new insights into the development, function, and evolution of hominoid teeth. This chapter is a brief review of recent advances. In essence, genetic analyses are changing our perspectives on the evolution of morphology, while improved studies of dental development and microstructure have yielded permanent markers of developmental history and microstructural differences of functional significance. More realistic perspectives on the physical properties of foods are yielding new functional interpretations of differences in tooth size. Finally, landmark-free analyses of tooth shape and wear are giving researchers the chance to actually monitor how teeth are used in living primates and by extrapolation in fossil primates too. Through techniques such as these will come a better understanding of the intricacies of dental function and a clearer picture of our past.

8.1 Introduction

People have been fascinated by the similarities between apes and humans ever since the first reports of apes filtered out of Africa. Linnaeus struggled to incorporate them into his System of Nature, but anatomical studies by Tyson (1699) and Huxley (1863), among others, forced the world to recognize the striking resemblance between African apes, in particular, and modern humans. Now, with the ape and human fossil record raising more questions, and technological advances generating new perspectives on morphology, it is time to take another look at the teeth of African apes, to see what we do, and do not know about them.

8.2 Dental development

Some of the most revolutionary discoveries in all of morphology have come in the areas of genetics and dental development, where subtle genetic changes have been
shown to have major impacts on morphology (Jernvall et al. 2000; Salazar-Ciudad et al. 2003). Thus far, studies have been restricted to laboratory animals such as rodents, but the potential impact on studies of morphological change in ape and human evolution are immense. With that point in mind, what do we know about dental development in the African apes? Methods of data collection have sometimes overlapped and sometimes varied dramatically, but some general trends are still evident in the literature.

Longitudinal data for nonhuman primates are rare, and this is certainly true for studies of dental development in apes, where methods of monitoring crown and root development have generally involved either detailed dissections or radiographs of different individuals of known ages (Zuckerman 1928; Dean and Wood 1981; Swindler 1985; Beynon et al. 1991; Conroy and Mahoney 1991; Winkler 1995; Kuykendall 1996) (although see Anemone et al. 1991 for an exception to this). Net results have included estimates of the timing of tooth calcification and emergence, and patterns of inter-tooth differences in those events. Analyses of chimpanzees have been far more common than those of other apes, due largely to the fact that chimps are often used in laboratory research. Ultimately, results are frequently compared with those for humans.

Estimates of the timing of tooth and root calcification have yielded not only some consistent patterns in apes and humans but also some patterns of differences between those groups. Specifically, dental development in apes is completed much more quickly than in humans, at approximately 11–12 years as compared with approximately 18–20 years (Nissen and Riesen 1964; Conroy and Mahoney 1991; Kuykendall 1996). Yet apes and humans have fairly similar patterns of cusp initiation and tooth mineralization (Swindler 1985), and their tooth crowns take similar times to develop (Dean and Wood 1981; Beynon et al. 1991). Two factors help clarify this apparent dilemma. First, there are significant differences in the degree of overlap between the development of certain teeth in apes and humans. For instance, in humans, each of the permanent molars completes its crown development before the next molar crown begins to develop. In apes, by contrast, there is a great deal of overlap in the timing of development of the molars (Reid et al. 1998; Dean 2000). A second factor helping to explain the faster completion of dental development in apes is a quicker rate of root growth after crown completion (Anemone et al. 1991, 1996; Simpson et al. 1992; Kuykendall 1996; Reid et al. 1998).

Finer-resolution differences in dental development may ultimately be discernible, both within and between species, e.g., differences between the sexes in some tooth mineralization stages (Kuykendall 1996) or differences in the formation time of specific molar cusps (Reid et al. 1998). However, additional resolution
will probably require further refinement of techniques (Winkler 1995; Beynon et al. 1998; Reid et al. 1998).

The net result of these developmental events is a complex pattern of sequences that can, by itself, be used to gain insights into primate life history (Smith 1991, 1994). Schultz (1960) recognized a basic distinction between relatively rapid- and slow-growing primates, with the former gaining all of their permanent molars before the eruption of more anterior permanent teeth. Hominoids are all relatively slow growing, but variations in their sequences of dental eruption (so-called tooth sequence polymorphisms) (Garn and Lewis 1963) can still yield insights. For instance, comparisons of humans and common chimpanzees show that humans have much greater variability in the eruption of their canines and lower central incisors, whereas chimpanzees show greater variability in the eruption of their second molars (Smith 1994). Still, if the actual timing of dental eruption can also be calculated (see Section 8.3.1 below), then the eruption of most pairs of teeth is highly correlated (Smith 1989), and the eruption of the first and last permanent teeth are, in turn, highly correlated with other life history variables.

### 8.3 Dental microstructure

Dental enamel is formed of hydroxyapatite crystals bundled together into prisms which are, in turn, often woven together in complex patterns, including radial enamel and decussating Hunter-Schreger bands (Martin et al. 1988; Koenigswald and Clemens 1992; Rensberger 1997; Maas and Dumont 1999). This inherent complexity has left researchers with a wealth of research possibilities, ranging from permanent markers of developmental history to structural anisotropy of functional significance.

### 8.3.1 Incremental microstructural features

Close examination of enamel prisms has revealed so-called “cross-striations” (periodic thickenings) laid down in a circadian fashion (Schour and Hoffman 1939; Massler and Schour 1946; Boyde 1964). This, coupled with surface markers known as perikymata, has allowed researchers to estimate the amount of time necessary for crown completion in modern hominoids (Dean and Wood 1981; Beynon et al. 1998; Reid et al. 1998; Shellis 1998; Dean 2000). But it has also given insights into tooth formation time and age at death in fossils (Bromage and Dean
1985; Beynon and Dean 1988; Dean et al. 1993), thereby suggesting that most of
the early hominids had an “apelike” pattern of dental development.

8.3.2 Enamel prism patterns

Before histological studies of the timing of dental development, researchers felt
that the shape of prisms in prepared tooth sections (prism-packing patterns)
could be used in phylogenetic studies, as certain patterns might be characteristic
of certain taxonomic groups (Shellis and Poole 1977; Gantt 1979, 1983). Howev-
er, subsequent work showed that results were often dependent on methods of
specimen preparation (Boyde et al. 1978; Vrba and Grine 1978). Detailed analyses
of enamel at controlled depths subsequently suggested that hominoids might
exhibit an unusual preponderance of “type 3” enamel (Boyde and Martin 1982;
Martin et al. 1988). However, more work is still necessary to document the range
of possibilities within and between large samples of teeth.

8.3.3 Enamel thickness

One obvious result of the complex process of tooth formation in most mammals
is an enamel cap covering the tooth crown—a cap that can vary rather dramati-
cally in thickness. Studies of molar enamel thickness in hominoids have gradually
progressed from simple linear measurements (Gantt 1977; Kay 1981) to more
complex measures designed to account for differences in body size (Martin 1983,
1985). Given the complexity of crown shape and development, it is perhaps no
wonder that there have been continuing discussions about proper methods of
analysis (Grine 1991, 2002, 2005; Macho and Thackeray 1992; Macho and Berner
1993; Macho 1994; Dumont 1995). However, if one were to use a summary
“measure” to characterize molar enamel thickness in hominoids, they could
probably be characterized as follows: Gorilla has relatively thin enamel, Pan and
Pongo range from thin to average thickness depending on tooth type, and Homo
has thick enamel (Shellis et al. 1998). These variations in enamel thickness are
probably due to differences in the duration of crown formation rather than the
rate of enamel production (Beynon et al. 1991).

Of course, along with differences in enamel thickness come questions of the
functional significance of those differences. Given the inherent complexity of
primate foods and diets, it is not surprising that the correlation between enamel
thickness and diet is not a perfect one (Maas and Dumont 1999), nor that the
physical properties of enamel may vary within the tooth crown and even between species (Cuy et al. 2002; Teaford et al. 2003). Over the past two decades, conventional wisdom has dictated that thick enamel-enabled primates to consume harder foods (Kay 1981; Dumont 1995). However, that perspective is changing largely because researchers are finally accepting the idea that enamel complexity, particularly prism decussation, serves as an admirable crack-stopping mechanism (Pfretzschner 1986; Koenigswald et al. 1987; Rensberger 1993, 2000; Maas and Dumont 1999). Thus it may be that prism decussation is a better correlate of hard-object feeding than is enamel thickness (Martin et al. 2003). In either case, we must remember that thick enamel may be an adaptive response to a variety of factors (Shellis et al. 1998).

8.4 Tooth size

Measurements of tooth size have been the focus of some classic studies of ape and human dentitions (Ashton and Zuckerman 1950; Schuman and Brace 1955; Garn et al. 1965; Pilbeam 1969; Mahler 1973; Johanson 1974; Swindler 1976). However, while some interspecific and intraspecific differences in tooth size are undoubtedly associated with differences in body size (Garn et al. 1968; Gingerich et al. 1982; Conroy 1987), the more intriguing trends are those differing from the standard assumption that larger animals simply have larger teeth.

For instance, investigators have long known that humans have relatively small canines compared to modern apes (Gregory 1922). However, among modern apes, chimpanzees and orangutans have relatively larger incisors than do gorillas and gibbons (Hylander 1975; Kay and Hylander 1978). As a result, differences in incisor size may reflect differences in the degree of incisor use in ingestion (Ungar 1996).

Intraspecific differences in canine size have also been used as indicators of sexual dimorphism, with gibbons and humans showing relatively little sexual dimorphism and chimps, bonobos, gorillas, and orangs all showing significantly higher sexual dimorphism (Ashton and Zuckerman 1950; Johanson 1974; Swindler 1976; Kinzey 1984). This has led to further inferences about differences in social behavior (Kelley 1986; Plavcan et al. 1995), with, for instance, species showing high degrees of sexual dimorphism also showing polygynous mating systems.

However, if analyses of tooth size are going to move from simple correlations (between morphological and behavioral differences) to explanations of causation, we need a better appreciation of the complexities of such relationships. In the 1970s, discussions began on the exact nature of the relationship between tooth
size and body size, with initial studies suggesting close ties between postcanine tooth area and body metabolism (Pilbeam and Gould 1974; Gould 1975), but subsequent analyses showing such a relationship to be grossly oversimplified (Kay 1975; Fortelius 1985). The confounding variables apparently come in two forms. First, analyses of modern primates have shown that the relationship between diet and food-processing is extremely complicated (Fortelius 1985; Lucas 2004). Thus, for example, both the rate of chewing and the physical response of food to chewing may influence the relationship between tooth size and diet. Second, analyses of fossil primates have shown that phylogenetic history may also complicate analyses of tooth size (Kay and Ungar 1997; Ungar 2002). For instance, Miocene apes generally have smaller incisors than modern apes.

So, is there a way to gain new insights from analyses of tooth size? Recent work by Lucas and coworkers (Lucas et al. 1986; Lucas 2004) suggests that there may be. In essence, anterior teeth and postcanine teeth are probably responding to different types of functional demands and thus need to be treated differently. Anterior tooth size is linked to the size of ingested particles, whereas posterior tooth size may depend on the deformability of the food, including a variety of properties like stickiness, particle shape, etc. Thus, perhaps relative differences in tooth size between incisors and molars in modern apes are giving us more subtle clues about the dietary differences between species. For instance, the fact that the gorilla has relatively small incisors compared to its molars may not simply indicate less reliance on the incisors in ingestion. It may also reflect the fact that the molars of gorillas often process relatively small food particles that are not very sticky (Lucas 2004). Clearly, however, more work is needed on the relationship between tooth size and the properties of foods.

8.5 Dental morphology and wear

As Aristotle noted nearly two-and-a-half millennia ago in De Generatione Animalium, tooth form reflects function. Studies of mammalian dental functional morphology do not date back quite that far, but they certainly boast a long and celebrated history nonetheless (Owen 1840; Gregory 1922). Early work on primate teeth, for example, suggested that their molars evolved to improve mechanical efficiency for particular masticatory movements (Crompton and Sita-Lumsden 1970; Kay and Hiiemae 1974). Primates that habitually crush foods have been noted to possess flat molar surfaces, whereas those that shear and slice have highly crested teeth (Rosenberger and Kinzey 1976; Seligsohn and Szalay 1978).

As in analyses of tooth size, dental functional morphology has recently begun to take more of a biomechanical perspective, focusing on relationships between
tooth shape and the strength, toughness, and deformability of foods (Strait 1993; Lucas and Teaford 1994; Spears and Crompton 1996; Yamashita 1998; Lucas 2004; Lucas et al. 2004). Primates that specialize on tough foods (those that are difficult to fracture), such as insect exoskeletons and mature leaves, generally have reciprocally concave, highly crested teeth for shearing and slicing. In contrast, those primates that prefer hard, brittle foods (those that resist initial puncture but are easy to fracture once a crack has started), such as many seeds, nuts, and palm fronds, tend to have rounder, flatter molar teeth for processing such items.

8.5.1 Quantifying functional aspects of tooth form

Of course, as we try to decipher more complicated relationships between dental form and function, quantitative approaches to characterizing tooth shape become essential. Researchers have thus developed several methods for characterizing tooth shape. The most popular of these is Kay’s (1978, 1984) shearing quotient (SQ) method. The lengths of mesiodistal crests are measured on unworn molars of several closely related species with similar diets. A least-squares regression line is fit to summed crest length and mesiodistal occlusal surface length in logarithmic space. SQs are computed as residuals or deviations from the regression line. This approach tracks diets of living apes fairly well, as the more folivorous siamang and gorilla have relatively longer shearing crests than do extant frugivorous hominoids (Kay 1977; Kay and Ungar 1997).

Spears and Crompton (1996) have suggested an alternative approach, measuring great ape cusp slopes from molar cross sections. They found that gorillas had high-angled occlusal surfaces, orangutans had gradually sloping surfaces, and chimpanzees had shallow “supporting” cusps, but steeper “guiding” slopes. These findings are taken to suggest that orangutans are adapted to reduce a hard/brittle diet, whereas gorillas can more efficiently fracture small food particles by shear. Chimpanzees, on the other hand, seem to be better suited to a diet with a wide range of mechanical properties. Such analyses probably give us more functional insights than do analyses of molar crown and cusp areas (Wood and Abbott 1983; Wood and Engelman 1988). However, as emphasized by Uchida (1998), the latter may still yield insights into a combination of genetic and ecological factors.

8.5.2 Tooth wear analyses

While studies of tooth morphology give glimpses of the complex relationship between primate tooth shape and diet, most such work has been limited to
unworn teeth. This is a major limitation because it leads to an incomplete picture of the form–function relationship. Wear is a normal phenomenon that begins as soon as a primate’s teeth come into occlusion. Thus, natural selection should also act on worn teeth, favoring morphologies that wear in a manner that keep them mechanically efficient for fracturing foods (Kay 1981, 1985; Teaford 1983; Teaford and Glander 1996; Ungar and Williamson 2000). In essence, since tooth wear occurs throughout an animal’s lifetime, we are missing a great deal of information if we exclude worn teeth in our analyses. Another limitation of studies that depends exclusively on unworn teeth is the lack of sufficient numbers of specimens for many (especially fossil) taxa. We find it remarkable, for example, that the entire published sample of early hominids from South Africa boasts less than 10 unworn M2s (the teeth most often used in functional studies).

The tooth wear of apes has occasionally been the focus of work in previous investigations. Early studies examined the degree of tooth wear in apes and humans in attempts to correlate tough, abrasive diets with the presence of increased tooth wear (Black 1902; Campbell 1925; Schultz 1935; Ashton and Zuckerman 1950; Welsch 1967). As noted by Wolpoff (1971), however, such interpretations are complicated by the complexities of diet and mastication, not to mention methodological difficulties associated with incorporating differences in dental eruption timing into such analyses.

Some investigators have made more detailed comparisons along the tooth row, noting, for example, that chimpanzees exhibit heavier incisor wear than do other apes (Ashton and Zuckerman 1950; Welsch 1967), again suggesting heavier incisor use in chimpanzees. More recently, Dean et al. (1992) have noted that later-erupting molars in chimpanzees and gorillas may actually show heavier wear than their predecessors, suggesting that occlusal loading is greatest on the last molar in the tooth row. It is perhaps no wonder then that molars in these species may also exhibit compensatory eruption as wear progresses (Dean et al. 1992) similar to that documented for some human populations (Whittaker et al. 1982, 1985).

With the advent of cineradiographic and electromyographic studies of mastication, studies of molar wear facets began to document subtle differences in jaw movement between primate species, with modern apes generally showing an increased emphasis on crushing and grinding as compared with some other catarrhines (Kay 1977; Maier and Schneck 1981). With these suggestions, analyses of tooth wear took another crucial step toward deciphering the relationship between dental form and function. If differences in the patterns of wear might indicate differences in jaw movement, could changes in tooth shape with wear yield further insights?
A look through the literature, however, reveals a striking paucity of such studies. Why? Quite simply, it is not easy to measure tooth shape on worn teeth. Traditional dental morphometrics depends on measuring distances between landmarks that are quickly obliterated by wear. Smith (1999) attempted to control for wear using a technique modified from Wood and coauthors (Wood et al. 1983). Molar occlusal views were captured on video and individual cusp areas were identified on a computer screen by mouse driven cursor. This allowed calculation of relative 2D (planometric) areas of cusps on unworn to moderately worn teeth (as long as cusp boundaries were identifiable). Smith’s results suggest that cusp proportions do indeed reflect diet to some degree, e.g., chimpanzees are linked with gibbons rather than gorillas.

Even this approach though is not ideal. First, specimens must still be sufficiently unworn to distinguish individual cusp boundaries, and these disappear pretty quickly, especially on thin enameled molars, such as those of chimpanzees and gorillas. More importantly, planometric area studies do not adequately characterize the third dimension of dental morphology. This is a problem because mastication occurs in a 3D environment, and two teeth with similar projected 2D areas may differ greatly in cusp relief.

The ability to collect elevation data is vital to studies of dental functional morphology. Cheek teeth have been known for the better part of a century to act as guides for jaw movements (Simpson 1933; Crompton and Sita-Lumsden 1970; Hiiemae and Kay 1972). Surface relief is critical to the angle of approach of mandibular and maxillary teeth as facets come into occlusion during mastication. This in turn determines the biomechanical efficiency with which items of given mechanical properties are fractured (e.g., whether foods are sheared or crushed).

### 8.5.3 Dental topographic analysis

What we need is a way to consider worn teeth in 3D studies of dental functional morphology. This is where dental topographic analysis comes in. Elevation data representing an occlusal surface are collected using a 3D scanner, and the tooth is modeled and analyzed using geographic information systems (GIS) software (Zuccotti et al. 1998; Jernvall and Selänne 1999) (Figure 8.1). Because dental topographic analysis does not depend on specific landmarks for measurement, it is equally useful for measuring unworn and worn teeth.

Some results for studies of living great apes are summarized and combined here to provide an example. Ungar and coworkers studied dental topography of...
Gorilla gorilla gorilla and Pan troglodytes troglodytes (M’Kirera and Ungar 2003; Ungar and M’Kirera 2003; Ungar and Taylor 2005). These taxa were chosen for analysis because of the modest degree to which they differ in the material properties of the foods they consume. At sites where the two taxa are sympatric, such as Lopé, Gabon, central African chimpanzees and western lowland gorillas overlap considerably in their diets, preferring soft, succulent fruits. The two taxa do differ though, especially at times of fruit scarcity. At such times, gorillas fall back more on tough, fibrous foods than do chimpanzees (Tutin et al. 1991; Remis 1997). Average annual food type proportions reported for central African common chimpanzees include about 70%–80% fruit flesh, as compared with 45%–55% fruit flesh for western lowland gorillas (Williamson et al. 1990; Kuroda 1992; Nishihara 1992; Tutin et al. 1997).

Dental topographic analysis of Pongo pygmaeus pygmaeus (Ungar and Taylor 2005) can add further insights into great ape molar form and function. The Bornean orangutan consumes an enormous variety of foods ranging from hard-husked, brittle nuts to soft fruits, to leaves, bark, and insects (MacKinnon 1977; Rodman 1977; Leighton 1993). While items consumed depend greatly on seasonal availability, average annual fruit to leaf proportions for Pongo pygmaeus pygmaeus are intermediate between those reported for Pan troglodytes troglodytes and Gorilla gorilla gorilla, with an average fruit percentage of about 55%–65% reported for the orangutans (MacKinnon 1977; Rodman 1977)—noting caveats concerning differences in data collection methods (Doran et al. 2002).

Data on average surface slope and occlusal relief are illustrated in Figure 8.2–8.4. These data are based on variably worn M2s of Pongo pygmaeus pygmaeus (n = 51), Pan troglodytes troglodytes (n = 54), and Gorilla gorilla gorilla (n = 47). Methods of data collection are presented in detail elsewhere (Ungar and Williamson 2000; M’Kirera and Ungar 2003; Ungar and M’Kirera 2003). Occlusal
surfaces were scanned as point clouds with lateral and vertical resolutions of 25.4 μm using a laser scanner. Resulting data files were opened as tables in ArcView 3.2 (ESRI Corp) GIS software, and digital elevation models were cropped to exclude areas below the lowest point of the occlusal basin. Average slope between adjacent points (surface slope) and the ratio of 3D to 2D planometric area (occlusal relief) were then recorded for each specimen.
Results are illustrated in Figures 8.2–8.4. The species overlapped in three wear stages (as defined in Ungar 2004). As expected, more worn molar surfaces of each taxon showed less occlusal relief and shallower slopes. At any given stage of wear, however, gorillas had the steepest slopes and most occlusal relief, followed by orangutans. Chimpanzees had the shallowest molar cusps and least occlusal relief.

This example suggests several things. First, tooth shape changes with wear. As teeth wear down, cusp slopes and occlusal relief both decline. Such changes likely affect functional efficiency. Further, apes with varying diets differ in the shapes of their teeth in ways that reflect the mechanical properties of foods that they eat. Species adapted to shearing and slicing tough leaves should have more occlusal relief and steeper sloped cusps than those adapted to crushing and grinding fruit. Cusp slope and occlusal relief values do mirror leaf-to-fruit ratios quite nicely for the great apes.

Another important point to come from this example is the notion that differences between species are of the same magnitude at different stages of wear. In fact, two factor ANOVA results show no significant interaction between species and wear stage for any of the variables examined (M’Kirera and Ungar 2003; Ungar and M’Kirera 2003; Ungar 2004; Ungar and Taylor 2005). This basically means that differences between species remain consistent through the wear sequence. In other words, we can compare chimps, gorillas, and orangutans at any given wear stage and get the same results. This is important because it means that species need not be represented by unworn teeth as long as there is a baseline of comparative data for specimens with similar degrees of wear. This will allow us to reconstruct the diets of a whole new assortment of fossil taxa that could not be analyzed in the past for lack of available methods.
8.5.4 Fallback foods and dental functional morphology

Another point emphasized by this work is the fact that gross differences in dental functional morphology need not relate to gross differences in preferred foods. Fecal studies for gorillas and chimpanzees at Lopé, for example, show 60%–80% plant species overlap (Williamson et al. 1990; Tutin and Fernandez 1993). These apes eat fruits much of the year but diverge at “crunch times” when preferred fruits are scarce. At such times, gorillas fallback more on leaves and other fibrous plant parts. The same is true for sympatric mountain gorillas and chimpanzees at the Bwindi Impenetrable National Park in Uganda (Stanford and Nkurunungi 2003). Differences between gorilla and chimpanzee occlusal morphology described here reflect fallback food choice more than everyday dietary preferences per se.

Apes have a penchant for succulent, sugar-rich foods—a legacy of the ancestral catarrhine dietary adaptation (Ross 2000; Ungar 2005). Differences in diet between catarrhines often rest largely with the seasonal shift to fallback foods taken when preferred resources are less available (Rogers et al. 1992; Lambert et al. 2004). In these cases, preferred resources are easy to digest, offer a low cost-benefit ratio, and may not result in selective pressures that would tax functional morphology. On the other hand, less desirable but seasonally critical fallback foods might require some morphological specialization (Robinson and Wilson 1998). This is not a new idea. Kinzey (1978), for example, noted that while *Callicebus moloch* and *C. torquatus* are both primarily frugivorous, the former has longer shearing crests for slicing leaves and the latter has larger talonid basins for crushing insect chitin. He reasoned that dental morphology therefore reflects adaptations not only to primary foods but also to less frequently eaten but still critical ones.

8.5.5 Function and phylogeny

As indicated earlier, in discussions of tooth size, another issue to consider when inferring dietary adaptations from morphology is the effect of phylogeny. Phylogenetic inertia or baggage plays an important role in how adaptations manifest themselves (Kay and Ungar 1997). We know, for example, that SQs track diet within cercopithecoids, hominoids, and platyrrhines—folivores have longer crests than frugivores within each of these higher-level taxa. On the other hand, cercopithecoids have relatively longer shearing crests than hominoids, and hominoids have relatively longer shearing crests than platyrrhines *independent of diet* (Kay and Covert 1984).
Because phylogeny determines the starting point for morphology, care must be taken when considering an extant baseline series to which an extinct species should be compared. One approach has involved considering ranges. For example, Early Miocene apes tend to have less well-developed shearing crests than do extant hominoids—though their ranges of SQ values are similar. It appears as if the extant hominoid range is upshifted relative to the Early Miocene ape range but reflects a comparable array of diets (Kay and Ungar 1997; Ungar et al. 2004). This can be confirmed by “anchoring” the range using independent data such as dental microwear patterning.

### 8.5.6 Dental microwear analyses

Fifty years ago, investigators realized they could gain insights into jaw movement and tooth use through light microscope analyses of wear patterns on teeth (Butler 1952; Dahlberg 1960; Dahlberg and Kinzey 1962). Subsequent work rekindled interest in the topic (Grine 1977; Rensberger 1978; Walker et al. 1978; Puech and Prone 1979; Ryan 1979) as many workers shifted to using the scanning electron microscope. Since then, analyses of modern and fossil material have yielded insights into dietary variations within and between species and also new perspectives on the evolution of tooth use and diet in animals ranging from dinosaurs to human ancestors (see Teaford 1994; Rose and Ungar 1998; Ungar 2002 for recent reviews).

The advantage of dental microwear analysis is that it provides evidence of what an animal was actually doing during its lifetime, not merely what it was capable of doing. Thus far, analyses of modern hominoids have generally been based on small samples (Gordon 1982, 1984; Teaford and Walker 1984; Teaford 1988; King et al. 1999), a factor which must be kept in mind when considering microwear data for animals, like chimpanzees and orangutans, with variable diets. Also, standard SEM analyses are proving to have a subjective component that may complicate comparisons of results between different investigators (Grine et al. 2002). Still, results to date suggest diets dominated by soft fruit for the chimpanzee; fruit with perhaps some hard objects for the orangutan; and tough, leafy vegetation for the mountain gorilla (Teaford and Walker 1984; Teaford 1988) (Figure 8.5). Interestingly, comparisons of results for lowland and mountain gorillas yield microwear differences suggestive of the dietary differences documented in the literature (Tutin and Fernandez 1993; Remis 1997), with lowland gorillas showing a higher incidence of pitting on their molars as compared with mountain gorillas (King et al. 1999) (Figure 8.6). Dental microwear analyses of *Pan paniscus* are just beginning.
As might be expected, dental microwear analyses of human ancestors have focused on whichever fossils are available. For the anterior teeth, qualitative studies have suggested similarities between early hominid incisor wear and that observed on modern primates that routinely employ a great deal of incisal preparation (Puech and Albertini 1984). Quantitative analyses of *Australopithecus afarensis* suggested similarities with lowland gorillas or perhaps savanna baboons (Ryan and Johanson 1989), while more detailed analyses of *Paranthropus robustus* and *A. africanus* (Ungar and Grine 1991) showed great variability within each species, but a greater density of features on the incisors of *A. africanus*, suggesting that *A. africanus* had a heavier emphasis on incisal preparation than in *P. robustus*.

In the molar region, qualitative analyses have raised many possibilities that have been often repeated in the literature. For instance, the robust australopithecines have been characterized as indistinguishable from modern chimpanzees or orangs (Walker 1981), perhaps with more abrasive molar wear than in the gracile australopithecines (Puech et al. 1985, 1986; Puech 1986b). Quantitative analyses have begun to refine these interpretations from a number of different sources.
perspectives. Studies of nonocclusal microwear have focused primarily on more recent European taxa, with initial analyses portraying the Neanderthals as more carnivorous than their immediate predecessors, or subsequent *Homo sapiens* (Lalueza Fox and Pérez-Pérez 1993; Lalueza et al. 1996). However, subsequent work has raised the possibility of sexual differences in diet in *Homo heidelbergensis* (Pérez-Pérez et al. 1999), and a more heterogeneous diet for the Neanderthals, with a shift in food processing in the Upper Paleolithic (Pérez-Pérez et al. 2003).

Quantitative analyses of fossil molar occlusal microwear began with Grine’s pioneering work on the South African australopithecines, where *Paranthropus robustus* was shown to exhibit more microwear and more pitting on its molars than did *A. africanus* (Grine 1981, 1986, 1987; Grine and Kay 1988). This leant
huge support to Robinson’s ideas of dietary differences among the australopithecines, with the so-called robust forms consuming harder foods that required more variable grinding movements in chewing. Recent work has taken analyses a step further by incorporating samples of australopithecines and early *Homo* from East and South Africa. Initial results give further credence to Ryan and Johanson’s (1989) idea of similarities between *Australopithecus afarensis* and lowland gorillas (Teaford et al. 2002a; Grine et al. 2006). Meanwhile, analyses of early *Homo* have begun to help sort through the variable assemblage that now encompasses that taxon, with *Homo erectus/ergaster* showing a higher incidence of pitting on its molars than that found in *Homo habilis* (Ungar et al. 2006), suggesting the consumption of tougher or harder food items by the former group.

As most studies of dental functional morphology ultimately hinge upon assumptions of the usefulness and selective advantage of the structures being measured, dental microwear analysis can certainly provide corroborative evidence for other hypotheses, for instance (as noted earlier) by “anchoring” analyses of dental morphological variation. Thus, dental microwear and molar shearing crest analyses for Eurasian fossil hominoids suggest that *Oreopithecus* was a folivore, *Ouranopithecus* was a hard-object feeder, and remaining forms such as *Dryopithecus* were soft-fruit eaters (Ungar et al. 2004). By contrast, molar shearing crest analyses of African Miocene hominoids suggest at first glance that all were either soft-fruit eaters or hard-object feeders. However, the microwear evidence suggests that molar shearing quotients for the African Miocene taxa are “downshifted” by about 50%, with *Rangwapithecus* as a folivore and the remaining taxa as soft-fruit eaters.

Obviously, knowing what we now do about primate diets, dietary categorizations such as these are gross oversimplifications. Either fallback foods or preferred foods may be of crucial importance for the survival and reproduction of individuals. Thus, either may be a selective force to be reckoned with in the evolution of morphological differences. So how can we tease them apart? The key lies in the collaborative use of as many lines of evidence as possible, on samples that are as large as possible (Teaford et al. 1996). For the fossil record, we have what we have, and we have to make do with it until more fossils are discovered. However, even now, the combination of dental microwear analyses and other morphological data has allowed researchers to document a gradual broadening of dietary capabilities in the earliest hominids (Teaford and Ungar 2000). This has then provided a backdrop against which to interpret the origin and early evolution of our genus (Teaford et al. 2002b), including, once again, insights into the distinction between dental *capabilities* and dental *use*, as the ability to process certain foods may well have been of critical importance in certain situations (e.g., meat-eating in *Homo erectus*) (Ungar et al. 2006).
The bottom line is that researchers have only begun to tap into the wealth of data, some from old sources and some from new. As noted above, new landmark-free analyses of morphology allow comparisons of tooth shape at any degree of wear, without the subjective identification of specific landmarks. Similarly, new objective methods of dental microwear analysis (Scott et al. 2005) allow the rapid characterization of microscopic wear surfaces for entire teeth in a matter of minutes. Through techniques such as these, we cannot help but gain a better grasp of variations in modern animals. With that will come a better understanding of the intricacies of dental function, and with that will come a clearer picture of our past.

Acknowledgments

We would like to thank Hartmut Rothe, Winfried Henke, and Ian Tattersall for inviting us to contribute to the Handbook of Paleoanthropology. Special thanks also go to Fred Grine, Chris Dean, Hartmut Rothe, and Winfried Henke for their comments during the preparation of this manuscript. As with any review paper, this one has benefited from the help of innumerable colleagues and curators from around the world and it would not have been possible without the support of the National Science Foundation whose help is gratefully acknowledged.

References

Black GV (1902) Dental anatomy, 4th edn. White, Philadelphia
Boyde A, Jones SJ, Reynolds PS (1978) Quantitative and qualitative studies of enamel etching
Dental adaptations of African apes

with acid and EDTA. Scanning Electron Microsc 1978 II: 991–1002
Campbell TD (1925) Dentition and palate of the Australian aboriginal. Hassell, Adelaide
Gantt DG (1977) Enamel of primate teeth: its thickness and structure with reference to functional and phyletic implications. Ph.D. dissertation, Washington University, St.Louis, Missouri, USA
Gregory WK (1922) The origin and evolution of human dentition. Williams & Wilkins, Baltimore, Maryland, USA
Huxley TH (1863) Evidence as to man’s place in nature. Williams & Norgate, London


Macho GA (1994) Variation in enamel thickness and cusp area within human maxillary molars and its bearing on scaling techniques used for studies of enamel thickness between species. Arch Oral Biol 39: 783–792


Mahler PE (1973) Metric variation in the Pongid dentition. Ph.D. dissertation, University of Michigan, Ann Arbor, Michigan, USA


dental development in the common chimpanzee, Pan troglodytes. J Hum Evol 35: 427–448


Tutin CEG, Fernandez M, Rogers ME, Williamson EA, McGrew WC (1991) Foraging profiles of sympatric lowland gorillas and...
Tyson E (1699) Orang-Outang: or the anatomy of a pygmy compared with that of a monkey, an Ape, and a Man Royal Society, London, England
Ungar PS (2005) Dental evidence for the diets of fossil primates from Rudabanya, northeastern Hungary with comments on extant primate analogs and “noncompetitive” sympatry. Palaeontographica Italica 90: 97–111

Abstract

The active intelligence of today’s primates flowered from trends that sculpted primate brain evolution across deep time: an increase in absolute brain size but a decrease in relative brain size (RBS, the ratio of brain size to body size) in bigger-bodied compared to smaller-bodied species (reflecting developmental scaling within species), increased RBS in highly “encephalized” species, and increased complexity of brain organization in conjunction with major adaptive shifts and selection for neurological specializations. Indices that quantify encephalization are discussed, as are developmental and physiological factors that constrain brain size. Data are provided which suggest that absolute and RBS increased steadily rather than erratically during the last 3 Myr of hominin evolution, and the “received wisdom” that human frontal lobes are disproportionately enlarged is questioned. Despite the enormous importance attributed to the evolution of primate brain size, the conviction remains that size alone is not enough to account for the observed diversity in primate behavior and that circuitry, neurochemistry, and subsystems (modules) were reorganized within brains to accommodate evolving behavioral repertoires (such as those entailed in language). Arguments about the relative evolutionary merits of brain size versus neurological reorganization are reviewed and, to some extent, reconciled.

9.1 Introduction

The mammalian order of primates is known for a variety of species that are lively, curious, social, and intelligent. Nonhuman primates are of special interest to people, not only because they are appealing and entertaining to watch but also because certain species (e.g., of macaques or baboons) are genetically close to humans, which makes them excellent animal models for medical research. As curious primates ourselves, we wonder about our evolutionary origins. One way to address this topic is to study and compare species from living primates that are thought to approximate broad stages (or grades) that occurred...
during some 65 Myr of primate evolution. Thus, one may compare particular anatomical structures or behaviors across appropriate representatives from the series prosimian → monkey → ape → human. When possible, such a comparative method should be supplemented with the direct method of studying fossil primates, which adds elements of specificity and time to the picture.

Within this broader context, we are also interested in the more specific question of how humans came to be, not only the largest-brained primate but also the most intelligent species on Earth. In order to address this question, one must study primate brain evolution. From our general understanding of primate evolution, we know that certain major adaptations occurred in some groups and that these changed and sculpted evolving brains during many millions of years. For example, the anthropoid ancestors of living monkeys, apes, and humans became diurnal, and this shift from night-living to day-living dramatically impacted the lives and nervous systems of their descendants forever after. Thus, brains of diurnal primates have relatively enhanced visual compared to olfactory “modules.” Another broad shift that greatly impacted the nervous system occurred when very early primates shifted from primarily ground living to living in trees. This newfound arboreal life underscored the adaptive value of keen vision with depth-perception and also led to improvements in sensory/motor coordination in conjunction with a variety of locomotor patterns that evolved in different arboreal species. More recently, some of these species shifted back to terrestrial living and this, too, left its imprint on their nervous systems.

As with many groups of mammals (Jerison 1973; Radinsky 1979), relative brain size (RBS) (the ratio of brain to body size) increased during the course of primate evolution. Some years ago, Radinsky (1979 p 24) noted that “elucidation of the factors responsible for the widespread evolutionary trend of increase in RBS in mammals, and for the extreme to which that trend was carried in humans, remains a fascinating unsolved problem.” Even before Radinsky’s observation, Jerison was pondering the laws that governed the evolutionary increase in brain size for various groups of primates (and other animals) and partitioning the respective total increases into two parts: those associated with allometric scaling expected for given body sizes, and any remaining increases (or decreases) in brain size, known as “residuals.” Other workers, most notably Holloway (1974, 1979), emphasized the evolutionary importance of neurological reorganization that alters the quantitative relationships between brain nuclei, fiber tracts, and neuroreceptor sites (Holloway et al. 2004), thus allowing for rewired and altered neurochemistry in brains of similar (or different) size. Although the debate about the respective importance of brain size versus neurological reorganization is a false dichotomy (Gould 2001) (both are important, of course), it continues today (Falk and Gibson 2001). Bringing welcome balance, Holloway et al. (2004) note
that the concept of reorganization in brain evolution is of less concern when one is examining broad genetic and evolutionary conservatism between large numbers of taxa (Jerison 1973; Finlay and Darlington 1995; Finlay et al. 2001; Kaskan and Finlay 2001), but more important when one attempts to explain species-specific differences in behavior (Preuss 2001; Holloway et al. 2004).

9.1.1 General methods for studying primate brain evolution

The direct method of studying fossilized braincases and casts of their interiors (endocranial casts or endocasts) is the bread and butter of paleoneurology (literally “old” neurology). Cranial capacities that approximate brain volume in cm$^3$ (and also brain mass in grams) may be measured from skulls by traditional methods such as filling braincases with mustard seed that is then measured in a graduated cylinder, or by obtaining volumes electronically from braincases (skulls) that have been subjected to three-dimensional computed tomography (3DCT). Indeed, because 3DCT is able to resolve small density differences, such as those between fossilized bone and attached rock matrix, it is particularly good for investigating fossils (Spoor et al. 2000), and has become useful as a noninvasive method for visualizing “virtual endocasts,” e.g., by flood-filling the virtual braincase (Falk 2004b). Although brain size is actually slightly smaller than cranial capacity because of the fluids, vessels, and brain coverings (meninges) that occupy the braincase along with brain tissue, the difference is insignificant compared to other sources of intraspecific variation in brain size (according to Hofman 1983, cranial capacity $= 1.05$ brain size), and the two variables are frequently used interchangeably. An advantage of using cranial capacities across the board in comparative studies is that, unlike actual brains, cranial capacities may readily be obtained from available skulls of fossil and extant primates.

Endocasts sometimes occur naturally under propitious geological conditions (such as those that exist in parts of South Africa) or, more often, are prepared artificially from skulls using liquid latex (see Falk 1986 for details). Over the past 20 years, the use of 3DCT data for reconstructing and measuring virtual endocasts has undergone numerous validation studies (Conroy and Vannier 1985; Conroy et al. 1990, 1998; Spoor et al. 2000) and is rapidly becoming a preferred method (Falk 2004b). Physical endocasts may be measured to determine cranial capacity, e.g., by displacing them in water (see Holloway et al. 2004 for details) and, as noted, the volumes of virtual endocasts may be measured electronically. Additionally, both kinds of endocasts (depending on their quality) may reveal positions of vessels and cranial nerves; details of suture closure, venous sinuses,
and emissary veins (foramina); information about cortical asymmetries including brain shape (petalia) patterns; and information about sulcal patterns. Curiously, the most detailed endocasts are produced from skulls of relatively young individuals within a species (Connolly 1950), and from skulls of smaller-brained species within a group of related species (Radinsky 1972). The former may relate to the timing of suture closure during development, while the latter may explain why some of the South African australopithecine natural endocasts reproduce a good bit of detail (Falk 1980a, b).

By comparison, those using the *indirect method* of comparing neuroanatomical structures among living species have a veritable arsenal of methods at their disposal. Specific cortical areas may be investigated using currently available histochemical and immunocytochemical techniques (Preuss 2001), in addition to relying on classic cytoarchitectural studies (Amunts et al. 1999). Questions can therefore be asked about the types, sizes, density, distribution, and connections pertaining to individual neurons, cell columns, or layers of the cerebral cortex (within and across particular regions). The comparative neuroscientist is able to ponder whether or not (and how) additional cortical areas have been “added” during primate evolution and the extent to which they might be associated with enlarged brains (Felleman and Van Essen 1991; Preuss and Goldman-Rakic 1991). Whereas CT is ideal for imaging fossil material, magnetic resonance imaging (MRI) is more suitable for imaging the soft-tissue structures that comparative neuroscientists study and may be performed noninvasively and in vivo. (Instead of relying on an X-ray source, MRI uses pulses of radiofrequency energy to map specimens that have been subjected to a strong magnetic field.) Even better, positron emission tomography (PET) and functional MRI (fMRI) are now commonly used to study functional processing in living human brains, and these techniques are beginning to be applied to nonhuman primates (Semendeferi 2001).

Although the increasingly sophisticated information gleaned from comparative brain studies is indispensable for interpreting paleoneurological data, the logistics of synthesizing findings from the *direct* and *indirect methods* for studying primate brain evolution remain tricky.

Deep disciplinary approaches, structural, functional, and developmental, have started to coalesce whereby brain structures can be seen developing and functioning through the many new noninvasive imaging techniques that are available today. All this allows better understanding of not only how the brain works in terms of movement and sensation but also how it functions during sleep, during preparation for action, during thinking, and during emotions. These lines of investigation, however exciting, and with such major implications for normal human brain function and in disease, employ more and more complex methods and reveal the workings of smaller and smaller brain components. Consequently,
the logistical problems of carrying out such studies in an evolutionary perspective and time-scale loom ever larger (Oxnard 2004 pp 1128–1129).

9.1.1.1 Quantifying primate brain size

Certain allometric factors govern the general external and internal morphology of primate brains. Larger primate (indeed, mammalian) brains are characterized by more convolutions (gyri and sulci) than smaller ones (Radinsky 1975), which appears to be a mechanism for maintaining the ratio of surface (cortex) area to brain volume as brains enlarge (Falk 1980b; Jerison 1982). (This is not to say that sulci and convolutions are never associated nonallometrically with specialized features. Sometimes they are [Falk 1982], e.g., brains of prehensile-tailed. New World monkeys have tail representations that are delimited by special sulci.) Neuronal density decreases with increased brain size, although mean neuronal size does not appear to scale allometrically with brain volume (Haug 1987). Compared to other mammals, the primate cerebral cortex is thicker and its layer IV is highly granulated (Haug 1987). The volume of gray matter is basically a linear function of brain volume, whereas the mass of interconnections that form the underlying white matter increases disproportionately with brain size (Ringo 1991; Hofman 2001). Curiously, women have relatively more gray matter than men (Haug 1987; see Falk 2001 for details regarding sexual dimorphism in primate brains).

Absolute brain size is hugely variable across living primates. Cranial capacities of living prosimians, monkeys, and gibbons overlap and together range between 1 and 205 cm³, which is separate from the great ape range of 275–752 cm³ (Falk 1986) (Figure 9.1). The human range is above that for great apes and extends from around ~1,100 to 1,700 cm³, excluding extreme outliers for purposes of comparison. But there is a problem here. The world’s smallest primate, the pygmy mouse lemur (Microcebus myoxinus), has a body weight of approximately 30 g (~1 oz), so how can we possibly compare its tiny brain size to those of larger primates such as the great apes? Clearly, a more meaningful parameter would be the ratio between brain size and body size, known as RBS. However, RBS is itself confounded by certain very powerful allometric scaling constraints that apply ontogenetically as individuals develop from smaller-bodied babies to adults (Passingham 1975b) and in interspecific comparisons of smaller-bodied with larger-bodied primates (Schultz 1956) (Figure 9.2). Allometric scaling is why human babies appear to have such relatively big heads (brains) compared to adults despite the fact that their absolute brain sizes are smaller, and it is why we should not be particularly impressed by the fact that little squirrel monkeys
have an average RBS of about 0.02, which is equivalent to that of humans (Falk and Dudek 1993).

In order to “subtract” the effects of allometric scaling, comparative studies of primate brain size have traditionally relied on quotients that express “residual” factors after the effects of body size have been removed from paleoneurological data (Falk 1980b). Thus, Bauchot and Stephan (1966, 1969) and Stephan (1972) developed the index of progression (IP) by using brain weight/body weight data from basal insectivores and calculating the regression equation:

$$\log h = 1.632 + 0.63 \log k$$

(9.1)

Figure 9.1
Ranges of cranial capacities in living primates, excluding far-reaching extremes in humans for comparative purposes (modified from Falk 1986)

![Figure 9.1](image)

Figure 9.2
Brain size and relative brain size (RBS) in humans and other primates. (a) Brain size growth in humans and chimpanzees. Brain growth in humans is at a higher rate after birth (b), which results in larger cranial capacities (and brain masses) for humans than chimpanzees at any given age (or body weight) (modified after Passingham 1975b). (b) Relative brain size (RBS, brain size/body size) plotted against body weight for humans, great apes, and Old World monkeys. The shapes of these curves are inverse to those above (in a) and, again, are stacked because of different rates of postnatal brain growth in the three groups. At any given body weight (and age), humans have RBS above those of apes, which are above those of monkeys. At smaller body weights (and therefore ages), primates have greater RBS (which is why human infants appear to have big heads) (modified from Schultz 1956). (c) Simple schematic that illustrates indices of relative brain size (1), the names of which vary (EQ, IP) with the reference group for the linear regression. When transformed to logarithms, brain size versus body size data (such as those in a) have a linear relationship...
(the straight line, or linear regression shown in c). \( P \) is the mean value for brain size predicted by the regression for a species at a given mean body weight; \( A \) is the actual mean value of brain size for that species. The index of RBS (or encephalization); \( i \) is the ratio of \( A \) to \( P \). The difference between \( P \) and \( A \), the residual \( r \), is the extra (or reduced) mean brain size that a species has compared to a species of similar mean body size in the reference group. The reference group (be it composed of mammals, insectivores, or just monkeys) is very important for interpreting indices.
where $h =$ brain weight, $k =$ body weight. From this equation, “basal” brain weight (BG) can be predicted for a given primate species by substituting its mean body weight into the equation. The ratio between actual mean brain weight of the species (progressive size = PrG) and the predicted “basal size” (BG) equals $IP$, the index of progression,

$$IP = \frac{PrG}{BG} \quad (9.2)$$

Jerison’s (1973) famous encephalization quotient (EQ) is similar, but uses brain weight/body weight data from living mammals rather than insectivores to establish the baseline regression and resulting classic formula:

$$EQ = \frac{E_i}{0.12P_i^{0.67}} \quad (9.3)$$

where $E_i =$ actual brain size, and $P_i =$ predicted brain size.

(Like other workers [Martin 1982, 1990], Jerison [2001] now uses a regression equation with an exponent of 0.75 instead of 0.67.) It should be noted, however, that the comparative results of EQ studies depend very much on the group selected for the baseline data (Holloway and Post 1982) and that there is an artifactual tendency for encephalization to be overestimated for smaller-bodied species but underestimated for larger ones (Radinsky 1982). One may also utilize similar regressions to estimate residual numbers of extra neurons (Jerison’s extra neuron index, $N_c$) or to determine how “encephalized” particular parts of the brain are. Toward these ends, Stephan et al.’s (1970) widely cited data for primate brains have been a gold mine for evolutionary studies.

### 9.2 The evolution of primate brain size

Compared to basal insectivores, primates evolved enlarged brain size/body size ratios (Radinsky 1975). Cranial capacity estimates for a dozen available Eocene and Oligocene prosimian skulls are all under 11 cm$^3$, and their EQs suggest that the Eocene lemuriforms *Smilodectes*, *Adapis*, and *Notharctus* were relatively smaller-brained that modern prosimians, and those for tarsiiformes (for which there is a brief series) increased through time (Radinsky 1975; Gurche 1982). By about 45 Ma, some prosimians appear to have RBS at the lower end of modern ranges (Radinsky 1975). Radinsky (1974) noted that an Oligocene anthropoid, *Aegyptopithecus*, had a cranial capacity of approximately 32 cm$^3$ but had nevertheless attained an anthropoid level of RBS. The Miocene hominoid *Proconsul* had a comparatively whopping endocranial volume of 167 cm$^3$ and an estimated body
weight of \(\sim 11\) kg, giving it a relatively bigger brain than modern monkeys of comparable body size (Walker et al. 1983). Over a quarter of a century ago, Radinsky (1974) summarized his findings from the “scanty” fossil record of primate brain evolution by observing that increased RBS dramatically distinguishes human brains from those of other primates, and he further suggested that this increase occurred relatively recently, beginning no more than 4–5 Ma (Radinsky 1975).

Although most modern anthropoids have brains that are relatively larger than those of modern prosimians (Bauchot and Stephan 1969), caution must be exercised when using EQs or similar indices to assess cognitive capacities. Despite the fact that EQs correlate to some degree with primate feeding behaviors (frugivorous primates are more encephalized than folivorous ones [Jerison 1973; Clutton-Brock and Harvey 1980; Milton 1988] and nonhuman primates that are omnivorous extractive foragers generally have higher Iqs than the others [Gibson 1986]), such indices fail to predict relative cognitive capacities. Gibson (2001) argues persuasively that, compared to monkeys, great apes possess greater mental constructional capacities and cognitive abilities in realms once thought to be uniquely human. Using a test for mental flexibility that separates apes from monkeys (the Transfer Index), she demonstrates that absolute brain size, body size, and extra neurons all correlate with performance, while EQ does not. Gibson therefore suggests that “the most practical measure for distinguishing intelligence and predicting the presence of humanlike mental skills in hominid fossils is absolute brain size” (Gibson 2001 p 92).

The relationship between primate brain size and cognition may also be explored by investigating the interaction between life history adaptations, brain growth, and cognitive levels (primate cognitive ecology [Garber 2004]). In one study, postnatal brain growth patterns were found to be highly variable among anthropoids (Leigh 2004). Leigh discerned two alternate life-history strategies that concern the metabolic costs of infant brain growth. In one, favored by Old World monkeys, relatively large-bodied mothers mature late and give birth to infants that require relatively little postnatal brain growth. This strategy requires high maternal metabolic investments during pregnancy. In the second strategy, exploited by tamarins, females mature especially early and produce offspring with brains that grow for a relatively long period of time during the postnatal period, which shifts some of their metabolic costs away from the mother and to others (including the offspring). Leigh notes that chimpanzees and humans are difficult to categorize in terms of these two strategies, and adds that differences in patterns of brain growth should be viewed as part of a more general complex of life-history traits rather than as direct pacesetters of life histories. Citing comparative studies on the cognitive abilities of squirrel monkeys, tamarins, and baboons,
Leigh concludes that life-history strategies may have coevolved with cognitive abilities in association with evolutionary changes in brain development.

Other studies investigate the perplexing question of how primates (including humans) were energetically able to grow relatively large brains that are metabolically “expensive” to maintain compared to the whole body. The maternal energy hypothesis (MEH) proposes that the mother’s relative basal metabolic rate (BMR) during an infant’s gestation determines its neonatal brain mass and that subsequent maternal investment while the infant is nutritionally dependent is also an important factor for developing big brains (Martin 1996). The MEH is sometimes contrasted with the expensive tissue hypothesis (ETH), which proposes that relatively encephalized primates are able to maintain their brain’s metabolic requirements because there has been an evolutionary trade-off in which brain tissue has increased at the expense (decrease in mass) of other metabolically expensive tissues such as guts, heart, liver, or kidney (Aiello and Wheeler 1995; Aiello et al. 2001). The two hypotheses should be viewed as complementary rather than contradictory because the MEH focuses on maternal energetics invested in offspring during gestation and lactation, while the ETH picks up from there by focusing on metabolic dynamics of brain growth and maintenance after weaning (Aiello et al. 2001). While the wider applicability of both hypotheses to mammals has been challenged by findings for bats (Jones and MacLarnon 2004), the recent trend toward studies that explore physiological and metabolic constraints on brain size and development is welcome and dovetails nicely with primate life-history studies. (Another constraint hypothesis about brain size evolution, the radiator hypothesis, concerns the evolution of vascular anatomy in response to brain temperature regulation combined with selection for bipedalism [Falk 1990, 2007].)

### 9.2.1 Parsing brain size evolution

But what are the possible neurological correlates of increased brain size during primate evolution? To address this, Finlay and colleagues concentrated on critical factors that drove mammalian (including primate) brain size, especially the manner and number of neurons generated during development (Finlay and Darlington 1995; Finlay et al. 2001). They found that the longer cytogenesis is prolonged for a given structure (based on timing of the peak in “neuronal birthdays”), the larger the structure will ultimately be. Since mammalian neurogenesis of brain parts proceeds uniformly (i.e., the order is conserved), “by far the most useful predictors of structure sizes are the sizes of other brain structures” (Finlay et al. 2001 p 268). (It should be noted, however, that olfactory bulbs
(and medulla) are an exception to this rule because they are smaller overall in anthropoids than prosimians and may map onto nocturnal versus diurnal niches [Barton et al. 1995].) A result of this regularity is that most parts of mammalian brains enlarged together, which led Finlay et al. to suggest that enlarged isocortices could have been by-products of structural developmental constraints (spandrels) that were only later co-opted for specific functions. The suggestion that the sizes of different brain structures is a consequence of overall brain size has perhaps received more controversy than it should have (Barton and Harvey 2000; Barton 2001; Oxnard 2004) because the Finlay et al. model, in fact, accommodates independent variation of individual brain parts that may be associated with specific behavioral advantages (e.g., foraging ability). Generally, this latter type of brain growth, which makes up the unaccounted-for variance in Finlay et al.’s (2001) model, underlies a small (but presumably evolutionarily crucial) variation of individual structure size on the order of two- to threefold.

9.2.2 Evolution of brain size in hominins

As noted, cranial capacities may be obtained for fossil hominins by measuring the endocasts (actual or virtual). Error may be introduced, however, because fossil endocasts are rarely whole and, thus, usually require partial reconstruction. Because morphological differences were found to distinguish the frontal lobes and temporal poles of robust and gracile (*Australopithecus africanus* australopithecines, new endocast reconstructions were provided for four *Paranthropus* specimens, which reduced the mean cranial capacity for the genus to the point where it approximated that of *A. africanus* (see Falk et al. 2000 for details). These new cranial capacities and others appear in Table 9.1 and Figure 9.3. (For more extensive data, the reader is referred to Appendix 1 of Holloway et al. 2004.)

A number of observations regarding the evolution of absolute brain size in hominins are suggested by Figure 9.3. Although brain size remained conservative during the evolution of *Paranthropus*, it increased in *Australopithecus* and between the latter and specimens that lived more recently (~1.7–1.9 Ma) in Africa and the Republic of Georgia. The overall morphology of these more recent specimens is transitional enough so that some workers place them in *Australopithecus* while others include them in early *Homo* (Wood and Collard 1999; Balter and Gibbons 2002). If, indeed, these specimens are transitional, then the received wisdom that brain size suddenly “took off” in the genus *Homo* around 2.0 Ma needs serious reevaluation (Falk et al. 2000; Falk 2004b). Thus, rather than there being a jump in cranial capacity in early *Homo*, cranial capacity may have begun increasing in the *Australopithecus* ancestors of *Homo* a million years ago
### Table 9.1

Cranial capacities for various adult hominins

<table>
<thead>
<tr>
<th>Species</th>
<th>Date (Ma)</th>
<th>Specimen</th>
<th>Adult cm³</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Australopithecus</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>A. afarensis</em></td>
<td>~3.2</td>
<td>AL 333-105</td>
<td>343</td>
<td>Falk (1987b)</td>
</tr>
<tr>
<td>–</td>
<td>~3.2</td>
<td>AL 162-28</td>
<td>375</td>
<td>Falk (1985)</td>
</tr>
<tr>
<td><em>A. africanus</em></td>
<td>~3.0</td>
<td>MLD 37/38</td>
<td>425</td>
<td>Conroy et al. (1990)</td>
</tr>
<tr>
<td>–</td>
<td>~2.75</td>
<td>Sts 60</td>
<td>400</td>
<td>Holloway et al. (2004)</td>
</tr>
<tr>
<td>–</td>
<td></td>
<td>Sts 71</td>
<td>428</td>
<td>Holloway et al. (2004)</td>
</tr>
<tr>
<td>–</td>
<td></td>
<td>Sts 5</td>
<td>485</td>
<td>Holloway et al. (2004)</td>
</tr>
<tr>
<td>–</td>
<td></td>
<td>Stw 505</td>
<td>515</td>
<td>Conroy et al. (1998)</td>
</tr>
<tr>
<td><em>Paranthropus</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>P. aethiopicus</em></td>
<td>~2.5</td>
<td>KNM-WT 17000</td>
<td>410</td>
<td>Walker et al. (1986)</td>
</tr>
<tr>
<td><em>P. boisei</em></td>
<td>~2.4</td>
<td>Omo L339y-6</td>
<td>427</td>
<td>Holloway et al. (2004)</td>
</tr>
<tr>
<td>–</td>
<td>~1.9</td>
<td>KNM-ER 23000</td>
<td>491</td>
<td>Brown et al. (1993)</td>
</tr>
<tr>
<td>–</td>
<td>~1.8</td>
<td>KNM-WT 17400</td>
<td>400</td>
<td>Holloway et al. (2004)</td>
</tr>
<tr>
<td>–</td>
<td>~1.8</td>
<td>OH 5</td>
<td>500</td>
<td>Falk at al. (2000)</td>
</tr>
<tr>
<td>–</td>
<td>~1.9</td>
<td>KNM-ER 407</td>
<td>438</td>
<td>Falk at al. (2000)</td>
</tr>
<tr>
<td>–</td>
<td>~1.7</td>
<td>KNM-ER 732</td>
<td>466</td>
<td>Falk at al. (2000)</td>
</tr>
<tr>
<td><em>P. robustus</em></td>
<td>~1.7</td>
<td>SK 1585</td>
<td>476</td>
<td>Falk at al. (2000)</td>
</tr>
<tr>
<td><em>Australopithecus/Homo?</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>–</td>
<td>~1.9</td>
<td>KNM-ER 1470</td>
<td>752</td>
<td>Holloway et al. (2004)</td>
</tr>
<tr>
<td>–</td>
<td>~1.75</td>
<td>D2700</td>
<td>600</td>
<td>Vekua et al. (2002)</td>
</tr>
<tr>
<td>–</td>
<td></td>
<td>D2282</td>
<td>650</td>
<td>Gabunia et al. (2000)</td>
</tr>
<tr>
<td>–</td>
<td></td>
<td>D2280</td>
<td>780</td>
<td>Gabunia et al. (2000)</td>
</tr>
<tr>
<td><em>Homo erectus</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Java (Sangiran)</td>
<td>~1.6</td>
<td>n = 6</td>
<td>Mean = 932</td>
<td>Holloway et al. (2004)</td>
</tr>
<tr>
<td>Africa</td>
<td>~1.5</td>
<td>KNM-WT 15000</td>
<td>909</td>
<td>Walker and Leakey (1993)</td>
</tr>
<tr>
<td>Java (Trinil)</td>
<td>~0.9</td>
<td>Trinil 2</td>
<td>940</td>
<td>Holloway et al. (2004)</td>
</tr>
<tr>
<td>China (Beijing)</td>
<td>~0.585</td>
<td>Skull D1</td>
<td>1020</td>
<td>Weidenreich (1943)</td>
</tr>
<tr>
<td>China (Beijing)</td>
<td>~0.423</td>
<td>n = 3</td>
<td>Mean = 1090</td>
<td>Weidenreich (1943)</td>
</tr>
<tr>
<td>Hexian</td>
<td>~0.412</td>
<td>–</td>
<td>1025</td>
<td>Wu et al. (2005)</td>
</tr>
<tr>
<td>Java (Solo)</td>
<td>~0.027</td>
<td>n = 6</td>
<td>Mean = 1149</td>
<td>Holloway et al. (2004)</td>
</tr>
<tr>
<td><em>Homo</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>European</td>
<td>~0.2</td>
<td>–</td>
<td>Mean = 1314</td>
<td>Hofman (1983)</td>
</tr>
<tr>
<td>Neandertals</td>
<td>~0.07</td>
<td>–</td>
<td>Mean = 1487</td>
<td>Hofman (1983)</td>
</tr>
<tr>
<td>European</td>
<td>~0.04</td>
<td>–</td>
<td>Mean = 1460</td>
<td>Hofman (1983)</td>
</tr>
<tr>
<td><em>H. sapiens</em></td>
<td>~0.01</td>
<td>–</td>
<td>Mean = 1330</td>
<td>Holloway et al. (2004)</td>
</tr>
<tr>
<td><em>H. floresiensis</em></td>
<td>~0.018</td>
<td>LB1</td>
<td>417</td>
<td>Falk et al. (2005)</td>
</tr>
</tbody>
</table>

Following Holloway et al. (2004), the chronological data are approximate middle values of the ranges for estimated dates (see also Falk et al. 2000). See Figure 9.3 for plots of data.
Cranial capacities of select hominins plotted against time; data from Table 9.1. (Above) plot includes capacities for robust australopithecines (Paranthropus). The trend for brain size increase appears flat until around 2.0 Ma, and then begins to increase in Homo. (Below) the same plot, but without the Paranthropus specimens (generally thought not to be ancestral to Homo). The trend toward brain size increase now appears to increase from before 3.0 Ma. Part of the reason for this is the recently described “transitional” specimens from Dmanisi, Republic of Georgia (listed under Australopithecus/Homo in Table 9.1). The earliest australopithecines and relatively recent LB1 (H. floresiensis) have brain sizes expected for apes of equivalent body sizes (%i = 1); H. erectus from Nariokotome (KNM-WT 15000) has a brain that is twice the size expected for similarly-sized apes (%i = 2); and contemporary H. sapiens’ mean brain size is three times that expected for apes of equivalent body size (%i = 3). This figure illustrates the trends for increasing brain size (vertical axis) and ongoing neurological reorganization (horizontal axis).
(Falk et al. 2000). With the redating of Java sites (Swisher et al. 1994; Huffman 2001) pushing certain cranial capacities further into the past, there is no longer the discontinuity in the trend for increasing cranial capacity (Falk 1987b, 1998) that once contributed to the suggestion that brain size evolution underwent “punctuated” events (Hofman 1983; Leigh 1992; Ruff et al. 1997). Rather, the recent discovery of LB1, the small-brained type specimen for Homo floresiensis (Brown et al. 2004; Morwood et al. 2004), lends an entirely new perspective to the study of hominin brain size evolution (Falk at al. 2005): from australopithecines through extant Homo, upward selection widened the range of brain size variation, while australopithecine-sized brains may have continued to provide the lower boundary (at least, until very recently). Thus, to some extent, Figure 9.3 encapsulates the interplay between selection for brain size (vertical vector) and selection for neurological reorganization (horizontal vector).

But what about the evolution of RBS in hominins? After all, LB1 was tiny, only about a meter in stature (Brown et al. 2004), which must certainly account for much of the recent variation in brain size. Although many workers have estimated EQs for fossil hominins, these estimates must be taken with a grain of salt because of the difficulty of determining surrogates for body mass. Without an associated skull, how does one identify the species of postcrania such as femurs that are often used to predict body mass or stature? Needless to say, the few known hominin partial skeletons are extremely important in this endeavor. Conservatively, we know exactly this much about encephalization in hominins: living people have brains and (separately) neocortices that are approximately three times as large as expected for nonhuman primates of the same body size (Stephan et al. 1970; Passingham 1973, 1975a; Passingham and Ettlinger 1974) and, surprisingly, this is true using regression equations based on all nonhuman primates, just monkeys and apes, or just apes (Stephan 1972; Falk 1980b). Turning to the hominin fossil record, there are two skeletons that provide important data. First, there is approximately 3.6 Ma. Although a definitive cranial capacity could not be obtained from this specimen, hominin cranial capacities of less than 400 cm³ were not uncommon at that time and place (Table 9.1), so it is safe to say that small-bodied australopithecines from Hadar, Ethiopia had ape-sized body masses that were probably associated with ape-sized brains (giving them a RBS index of $i = 1$). Fast forwarding to ~1.5 Ma, the H. erectus skeleton from Nariokotome, Kenya (KNM-WT 15000) paints quite a different picture. By the time he reached adulthood, it was projected that this “lad” would have reached a stature of over 6 ft and a cranial capacity of 909 cm³ (Walker and Leakey 1993). That capacity is twice the means for both A. africanus and Paranthropus (Table 9.1), and roughly twice the means for living great apes (490 cm³ for gorillas, 375 cm³ for
common chimpanzees and for orangutans [Falk 2000b p 312]). It is also 2/3 of 1364 cm$^3$, which is very close to the oft-cited world mean for contemporary $H.\ sapiens$ of 1350 cm$^3$. It therefore looks as if African $H.\ erectus$ that lived $\sim$1.5 Ma may have had a brain mass that was twice the size predicted for a living nonhuman primate of equivalent body mass ($i = 2$) or, put another way, that $H.\ erectus$ was two-thirds as encephalized as $H.\ sapiens$. These few data provide nice 1–2–3 estimates for indices of RBS at $\sim$3.0 Ma, $\sim$1.5 Ma, and today (Figure 9.3). Beyond this, conjecture about the past evolution of hominin encephalization remains just that.

One can, however, make reasoned conjectures about future brain size evolution. In a fascinating paper, Hofman (2001) applies the design principles and operational modes (including energetic and neural processing constraints) that underlie information processing in primate brains to the task of modeling the limits of future brain size evolution in $H.\ sapiens$. His model predicts that, as brain size increases beyond a certain critical point, subcortical volume (cerebellum, brain stem, diencephalon, etc.) would decrease in conjunction with increasing white matter. The net result would be that hominins with brains enlarged beyond that critical point would have a declining capability for neuronal integration despite an increased number of neurons. The critical point is $\sim$3500 cm$^3$, beyond which “any further step in the evolution of intelligence will then have to take place outside our nervous system, in a technological world where the selection mechanisms and forces are radically different from those operating in nature” (Hofman 2001 p 125). Although Hofman’s model does not incorporate anatomical constraints that govern head size and parturition, perhaps the technological world he envisions will, indeed, make it possible for women to bear (presumably) bigger-brained neonates.

### 9.3 The evolution of neurological reorganization

Despite the enormous energy that paleoneurologists have devoted to studying primate brain size evolution, there remains a conviction that size alone is not enough to account for the observed diversity in primate behavior, and that circuitry, neurochemistry, and subsystems (modules) must have become reorganized within brains to accommodate evolving behavioral repertoires (Preuss 2001; Holloway et al. 2004). Preuss, in fact, goes so far as to suggest that “the cortex is a veritable hotbed of evolutionary reorganization” (2001 p 140). Although reorganization was undoubtedly important, deciphering the details of internal brain evolution is much more difficult than studying the gross phenomenon of brain size. Nevertheless, information yielded by both direct and indirect methods...
sheds some light on at least the broad aspects of neurological reorganization that occurred during primate evolution.

Comparisons of brains of basal insectivores and living primates suggest early evolutionary trends in primates that included not only the larger brain size/body size ratios noted above but also relatively enlarged neocortices for brain size, a decrease in the relative size of the olfactory bulbs, an increase in the amount of visual cortex, and development of a central sulcus in anthropoids rather than the coronal sulcus seen in prosimians (Radinsky 1975). At histological levels, layer 4 of the posterior cingulate cortex appears to be less densely packed with small cells in prosimians than anthropoids (Zilles et al. 1986), a finding that has now been extended to include much of the parietal and temporal cortices (Preuss and Goldman-Rakic 1991). The fossil record of prosimian endocasts helps pin down the approximate dates when some of these primate specializations occurred (Radinsky 1974, 1975; Gurche 1982). Thus, visual and temporal cortices had expanded to comparable modern levels in some ancestral tarsiiform and lemuri-form primates by ~55 Ma (Early Eocene), but frontal lobes were still relatively small except in the line leading to *Adapis* (Radinsky 1975; Gurche 1982). Analysis of the fossil record of anthropoid endocasts, particularly partial endocasts of *Aegyptopithecus*, reveals that by ~25–30 Ma (Oligocene), olfactory lobes had reduced and visual cortices had expanded compared to prosimians. Although its frontal lobes appeared to be small compared to modern anthropoids, *Aegyptopithecus* had an anthropoid-like central sulcus instead of a longitudinally oriented fissure, the coronal sulcus, which separates head from forelimb representations in primary somatosensory cortices of prosimians (Radinsky 1975). The oldest record of an anthropoid endocast of modern appearance is that of ~18 Ma *Proconsul* (Falk 1983; Walker et al. 1983). Regarding neurological reorganization during primate evolution, Radinsky (1974 p 25) summarized:

▶ Since *Aegyptopithecus*, *Dolichocebus*, and *Apidium* are among the oldest known pongids, ceboids, and cercopithecoids, respectively, it is likely that elaborations of visual abilities and reduction of olfaction were among the features involved in the initial emergence of higher primates from prosimians. It is interesting that those same features, although not as extensively developed, appear to have been among the key adaptive features at the base of the great Eocene prosimian radiations.

A comparative study of endocasts from extant New and Old World monkeys describes various cortical specializations that were independently evolved in both groups as well as similarities that were retained from a common ancestor (Falk 1981). Within Old World monkeys, cercopithecine sulcal patterns appear to be more derived than colobines as manifested in relative expansion of prefrontal,
and inferior temporal integration cortices (Falk 1978). Radinsky (1974) showed that a cercopithecoid endocast from *Mesopithecus*, dated to ~9 Ma, exhibits the typical colobine pattern and is similar to the brain of ~6 Ma *Libypithecus*. Thus, the modern colobine sulcal pattern, which appears to represent the more primitive condition, had occurred by at least 9 Ma. He also noted that the derived cercopithecine sulcal pattern had appeared by ~2 Ma in *Paradolichopithecus*.

The addition of new cortical areas may have provided an opportunity for the evolution of new behavioral capacities (Kaas 1987, 1995; Allman 1990, 1977; Felleman and Van Essen 1991; Preuss and Goldman-Rakic 1991). To date, primates are known to possess 50–100 cortical areas, and it has been hypothesized that many of these may be higher-order areas that are unique such as dorsolateral prefrontal, posterior parietal, and inferotemporal cortices (Preuss 2001). Preuss also notes that higher-order association regions of primates are strongly connected with each other and these regions are all connected with a prominent thalamic structure, the medial pulvinar, which has no obvious counterpart in other mammals. He further suggests, “not only do primates possess primate-specific higher-order cortical territories, but these territories form a distinctive connectional system” (Preuss 2001 p 153). The suggestion that new cortical areas constitute a natural by-product of increasing brain size is consistent with Ringo’s (1991) mechanistic observation that enlarging brains would become swamped with white matter without neurological reorganization that increased the number of local (as opposed to longer corticocortical) connections and therefore areas (Hofman 2001).

Relatively recent comparative work also suggests that the cerebellum, long known to be important for motor coordination and now thought to contribute to higher cognitive functions in humans (Fiez 1996; Muller et al. 1998), underwent neurological reorganization during primate evolution. Thus, the lateral cerebellar system is relatively large in chimpanzees and gibbons, while a central nucleus (the dentate nucleus, the output of which influences the cerebral motor cortex) is larger in humans than apes (Matano et al. 1985; Matano and Hirasaki 1997). This is particularly interesting in light of the fact that the human cerebellum appears to be smaller than expected for an ape brain of human size (Semendeferi and Damasio 2000).

It is important to keep in mind that a part of the brain does not need to be “new” or grossly enlarged for reorganization to occur. For example, Armstrong et al. (1987) investigated which thalamic nuclei changed in volume relative to the rest of the thalamus and found that, after controlling for the size of the brain, anthropoids that lived in single-male societies had more anterior principal thalamic neurons than primates that lived in multimale societies. Since limbic structures are known to be important for social life, it is not surprising that the
sizes and reorganizations of limbic structures may link more than those of other structures to specific behaviors and niches (e.g., the relationship of olfactory bulbs with nocturnal and diurnal niches) (Finlay et al. 2001). (Along somewhat related lines but focusing on gross brain size rather than reorganization, the “social brain” [or Machiavellian Intelligence] hypothesis incorporates data showing that neocortical size correlates with social group size and proposes that large primate brains evolved in response to living in complexly bonded social groups [Falk and Dudek 1993; Dunbar 1998, 2003; Byrne 2000].)

While complex social life may, indeed, have contributed directly or indirectly to selection for large primate brains, partitioning the types of internal reorganization that characterize different groups hones in on other aspects of lifestyle. In an important follow-up to Finlay et al.’s research, de Winter and Oxnard (2001) performed similar multivariate analyses on a greatly enlarged data set that confirmed earlier findings (Finlay and Darlington 1995) and extended earlier multivariate analyses to include a series of brain-part ratios that partly reflected input/output relationships within the brain. Rather than grouping primates according to phylogenetic relationships, however, the groups that emerged from the comparisons were based on similar lifestyles, such as lower-limb dominated lifestyles that involve much leaping (tarsiers, indriids, galagos, mouse lemurs), and four-limb dominated lifestyles (some strepsirrhines, New and Old World monkeys) (de Winter and Oxnard 2001; Oxnard 2004). Genera with upper-limb-dominated lifestyles involving hang-feeding in arboreal habitats and escaping by upper-limb acrobatics (Ateles, Lagothrix, Hylobates, Pan, and Gorilla) also emerged as a cluster.

The brain organization that is involved in the trend along the axis toward the forelimb dominant species is increasing expansion of the neocortex, striatum, cerebellum, and diencephalon relative to medulla. This particular pattern of brain organization could involve brain functions based on expansion of higher levels of voluntary sensory and motor control. In turn, they could relate to a trend toward creatures with greater degrees of complex voluntary behavior and increased capacity to plan strategically and to control complex motor actions (Oxnard 2004 p 1147).

Significantly, similar multivariate analyses separate humans from chimpanzees to a degree that rivals the extent of separation within all Old World monkeys and apes, which is not only contrary to the much-cited close genetic relationship between Pan and Homo but also implies that the internal organization of the human brain is quantitatively different from any other living primate (Oxnard 2004). Further, the differences between chimpanzees and humans are not related to brain size alone and may relate to the existence of internal functional interactions, loops, or modules (Oxnard 2004).
9.3.1 Neurological reorganization in hominins

In light of the theoretical emphasis neurological reorganization has been given in the literature, surprisingly little precise information is available about its nature during hominin evolution. However, recent work by a few workers provides a glimpse of what might have happened. Contrary to earlier notions about “mosaic evolution,” the research of Finlay, Oxnard, and colleagues discussed above suggests that major steps in neurological reorganization (i.e., as opposed to, say, fine tuning of individual nuclei) rarely, if ever, entailed isolated structures within the brain but, instead, were probably distributed across multiple structures (or modules) within the brain. This hypothesis is concordant with functional imaging studies that indicate higher-order cognitive tasks engage numerous cortical areas that are dispersed across the cortical mantle (Frackowiak et al. 1997). For this reason, I remain respectfully skeptical about the suggestion (based on controversial identifications of the lunate sulcus) that early australopithecines with otherwise apelike cortical morphologies were reorganized in posterior parietal association cortices (Holloway et al. 2004). That said, I agree with my colleagues that dramatic neurological reorganization can occur separately from brain enlargement but hypothesize that it is likely to be manifested more globally.

Recent research of Semendeferi and colleagues (Semendeferi et al. 1997; Semendeferi and Damasio 2000; Semendeferi 2001) sheds light on neurological reorganization in hominins at the level of large sectors including whole lobes. Semendeferi and Damasio (2000) obtained MRI scans of brains from nearly 30 living humans and apes, processed the data to obtain volumes of the various lobes, and performed comparative statistical analyses of the absolute and relative volumes of each lobe. Although the overall relative sizes of the lobes of the brain changed little after the phylogenetic split of hominins from great apes, this study revealed that the temporal lobe (involved in recognition and memory) may have differentially enlarged during hominin evolution, while the human cerebellum is significantly smaller than expected from allometric predictions. The insula (which processes autonomic functions, internal stimuli, taste, and speech articulation [Dronkers 1996]) may also be somewhat enlarged in humans. Contrary to Semendeferi’s (2001) finding of no increase beyond allometric expectations for the large parieto-occipital sector of the human brain, those of another study (utilizing geometric morphometrics) suggest that modern humans are characterized by relatively great development of the parietal lobes (Bruner 2004), a conclusion that awaits further confirmation.

Equally important, by analyzing 3D-MR reconstructions of brains from living apes and humans (Semendeferi et al. 1997) in conjunction with comparative histological sections from postmortem specimens, Semendeferi and her
colleagues have helped dispel old myths (indeed, some might even say “received wisdom”) about the evolution of human frontal lobes. Until recently, many believed that higher cognitive abilities in humans evolved in conjunction with differentially enlarged frontal lobes. Semendeferi’s comparative imaging work dispelled this notion, however, by quantifying the allometric nature of human frontal lobe enlargement (i.e., they are the size one would expect in ape brains enlarged to the size of human brains).

Turning to the important question of neurological reorganization within larger sectors, comparative cytoarchitectonic studies suggest that human frontal lobe evolution entailed internal rewiring and enlargement in some areas (e.g., Brodmann’s area 10, Semendeferi et al. 2001, 2002) and a decrease in others (Brodmann’s area 13, Semendeferi et al. 1998) rather than an increase in overall frontal lobe size. It was therefore concluded that area 13 of the posterior orbito-frontal cortex, a part of the limbic system that is involved in emotional reactions to social stimuli, is a conserved feature in brain evolution, whereas the relative size of area 10 that forms the frontal pole in ape and human brains and contributes to planning and the undertaking of initiatives did increase during hominin evolution. A remarkable increase in the proportion of white matter volume of the human precentral cortex was also found (Semendeferi et al. 1997), which again speaks to the fact that human frontal lobes are better wired rather than relatively larger than those of their ape cousins.

Recent work also suggests that the visual system was reorganized during human evolution (Preuss et al. 1999), which surprised even the investigators because “it is axiomatic among neuroscientists and psychologists that the visual abilities of humans and monkeys are virtually identical” (Preuss 2001 p 156). Specifically, the authors report histological evidence suggesting that the human primary visual area differs from that of apes and monkeys in the way that information is segregated from layers of the lateral geniculate nucleus. Interestingly, they suggest that humans have enhanced capacities for analyzing moving stimuli and speculate that these changes may have occurred in response to the challenge of visually decoding rapid mouth movements entailed in speech and its accompanying manual gestures (Preuss 2001).

9.3.1.1 Neurological reorganization related to language, handedness, and music

It is tempting to hypothesize that the expansion of the human cortex was accompanied by the addition of new areas and that the classic language areas in the left hemisphere (Broca’s speech area [Brodmann’s areas 44 and 45] and
Wernicke’s language receptive area [Brodmann’s areas 21, 22 plus, when defined more broadly, 37, 39, 40]) are neomorphic structures (Preuss 2001). However, Preuss notes that “at the present time, there is no good evidence that humans possess species-specific cortical areas” (Preuss 2001 p 155). Indeed, cytoarchitectonic studies on macaques suggest that the inferior limb of the arcuate sulcus contains homologs of areas 44 and 45 (Galaburda and Pandya 1982; Deacon 1992; Preuss 2000), and homologs of posterior language areas (Wernicke’s area) have been identified in the macaque superior temporal and inferior parietal lobes (Galaburda and Pandya 1982; Preuss 2000) (Figure 9.4). Simple movements of the mouth and hands activate ventral premotor cortex in monkeys, as they do its likely homolog, Broca’s area, in humans (Petersen et al. 1988; Colebatch et al. 1991; Gallese et al. 1996; Rizzolatti et al. 1996), and these “mirror neurons” also

**Figure 9.4**

Gross language areas in humans and their proposed homologs in macaques and common chimpanzees. In the left hemispheres of humans, Brodmann’s areas 45 (pars triangularis) and 44 comprise Broca’s speech area, while areas Tpt (temporoparietal), PT (planum temporale, buried within depths of Sylvian fissure), and Brodmann’s area 40 are parts of Wernicke’s receptive language area. Human area 40, macaque area 7b, and chimpanzee area PF/PG are proposed homologs, as are human and macaque areas Tpt and chimpanzee area TA. The proposed homologs are based on cytoarchitectonic and functional similarities and should be viewed as tentative. Data from Preuss (2000), Amunts et al. (1999), Gannon et al. (1998), Aboitiz and Ricardo (1997), Galaburda and Pandya (1982), Crosby et al. (1962), Jackson et al. (1969), Bailey et al. (1950), von Bonin (1949)
discharge when similar actions are observed in others (Rizzolatti et al. 1996). Because of their discovery in human and nonhuman primates, mirror neurons are hypothesized to be part of an action-perception network that facilitates gestural (manual and orofacial) communication in apes and humans as well as linguistic communication in the latter (Falk 2004c, d). From a functional perspective, it is also interesting that, like humans, macaques are thought to be left-hemisphere dominant for processing certain socially meaningful (as opposed to neutral) vocalizations (Petersen et al. 1978, 1984; Heffner and Heffner 1984, 1986).

Paleoneurologists have long speculated about whether a chimpanzee-like frontal lobe in early hominins could have given rise to a humanlike Broca’s area, but these efforts have been hampered by a lack of consensus about the identities of homologous sulci and gyri in great apes and humans, which were traditionally proposed mainly on the basis of relative positions of sulci rather than on cytoarchitectonic grounds (Connolly 1950). Unlike frontal lobes of humans, a fronto-orbital sulcus (fo) of chimpanzees typically incises the lateral border of the dorsal frontal lobe and extends onto its orbital surface where it courses caudally to the temporal pole (Connolly 1950). The bulge delimited by fo, or so-called orbital cap, represents Brodmann’s area 44 (Bailey 1948; Bailey et al. 1950; Connolly 1950; Jackson et al. 1969) (Figure 9.4) and to varying degrees the addition of part of area 45 (Sherwood et al. 2003) in chimpanzees. Sherwood et al. (2003) explored the relationship of sulci to cytoarchitectural areas 44 and 45 in brains from five adult chimpanzees and found that, just as the border between cytoarchitectonic areas 44 and 45 of humans is not always defined by sulci (Amunts et al. 1999), the border between the two areas in chimpanzees does not always coincide with the surface of the fronto-orbital sulcus. Rather, inter-subject variability was high and area 45 tended to spill over caudally into the presumed domain of area 44 in both species. It is also important to stress that the similar bulge that appears at the level of the temporal pole in humans, the orbital cap (or so-called “Broca’s cap”), is not homologous to that of chimpanzees because it contains areas 45 and 47 rather than the areas located in the chimpanzee cap, namely, area 44 (Connolly 1950) and (sometimes) 45 (Sherwood et al. 2003). Although it has recently been suggested that area 44 is larger in the left than the right hemisphere of chimpanzees (Cantalupo and Hopkins 2001) as is the case for humans (Amunts et al. 1999), for methodological reasons the jury is still out on whether or not the homolog of Broca’s area in great apes exhibits humanlike asymmetry (Sherwood et al. 2003).

There is more agreement about asymmetry in the chimpanzee homolog of at least part of Wernicke’s area. Gannon et al. (1998, 2001) investigated the homolog of the planum temporale (PT) in 18 chimpanzee brains and determined that the
left PT was significantly larger in 17 of the 18 brains (94%). This region is a component of Wernicke’s area in the left hemisphere of humans, in whom it manifests a similar anatomical pattern and left hemisphere size predominance. The authors concluded that human language may have been founded on this basal anatomic substrate and that it may have been lateralized to the left hemisphere in the common ancestor of chimpanzees and humans millions of years ago (Gannon et al. 1998).

Thus, more than a century after Broca’s area was identified, it is recognized that it has certain nonlinguistic functions and that the act of speech activates wider areas of the cerebral cortex. Nevertheless, the importance of this area for speech and Wernicke’s area for human language reception cannot be denied, and the evolutionary details of their coordinated neurological reorganization (including with other parts of the brain) remain open to investigation (Sherwood et al. 2003; Holloway et al. 2004).

One may, however, engage in reasoned speculation about the evolution of a suite of unique behaviors in hominins and their underlying interconnected and reorganized neurological structures. We know, for example, that people are more neurologically lateralized than other primates and that certain cortical asymmetries underpin behaviors that are unique to the human primate (Falk 1987a), such as the universally high frequency of right-handedness, symbolic language, and humanlike creative abilities related to music, art, and technology (Falk 2000a, 2004a). One may explore the evolution of brain lateralization by studying shape asymmetries in endocasts of fossil hominins (Holloway et al. 2004), since in living people these petalias (which exist to a lesser extent in nonhuman primates [LeMay et al. 1982]) are statistically associated with handedness patterns and sex (LeMay 1977; Bear et al. 1986). We also know that men and women differ in the anatomies of their brains, and that these differences are hypothesized to have evolved as correlates of different reproductive strategies (Falk 1997, 2001; Falk et al. 1999). Although a review of the literature on primate brain lateralization is beyond the scope of this chapter, it is worth noting that Hofman’s (2001) exploration of design principles that govern the evolution of large brains led him to conclude that large brains tend to increase the number of distinct cortical areas in order to maintain processing capacity, and that this may be related to the high degree of brain lateralization in humans.

Large-brained species may develop some degree of brain lateralization as a direct consequence of size. If there is evolutionary pressure on certain functions that require a high degree of local processing and sequential control, such as linguistic communication in human brains, these will have a strong tendency to develop in one hemisphere (Aboitiz 1996; Hofman 2001 p 123).
9.4 Conclusions

Primate nervous systems became more variable over the course of evolution. During the Eocene, brain sizes were all small. Today, there are still small-brained species, but also larger-brained ones due to a widening range of variation as the Cenozoic progressed. The same can be said for RBS. Over thirty years ago, Radinsky (1974) pointed out that elaboration of visual abilities and reduction of olfaction were among the features involved at the base of prosimian radiations and, again, in the later emergence of higher primates from prosimian stock. The broad visual and limbic systems that subserve these features were (and are) extremely important for primate species-specific communication. Over time, the various neurological components of these systems became variably elaborated and reorganized within different groups. Preuss’ (2001) suggestion that the “surprisingly” reorganized human visual system may have evolved in response to the challenge of visually decoding rapid mouth movements entailed in speech and its accompanying manual gestures underscores the ongoing continuity of adaptations that occurred extremely early in primate evolution. Semendeferi’s (2001) seminal work on hominoid prefrontal cortices (Brodmann’s areas 10 and 13) illustrates that executive parts of the cerebral cortex eventually got into the act and were also subjected to evolutionary reorganization (Semendeferi et al. 2001).

The arguments about the relative evolutionary merits of brain size versus neurological reorganization are unnecessary (Gould 2001). The suggestion by Finlay and colleagues (Finlay and Darlington 1995; Finlay et al. 2001; Kaskan and Finlay 2001) that the sizes of different brain structures is a consequence of overall brain size, not only in primates but also in other mammals, is an important contribution to our understanding of ontogenetic brain development and brain evolution. What has sometimes been lost is that Finlay’s model leaves room for evolution of the kinds of neurological specializations that interest paleoneurologists. Oxnard and de Winter’s models for parsing brain size evolution are no less elegant and shed light on the evolution of broad (but presumably intertwined) subsections of the nervous system that subserve very different lifestyles, separate from phylogenetic considerations (de Winter and Oxnard 2001; Oxnard 2004). These findings extend, rather than contradict, those of Finlay and colleagues. Add to the mix, neurological reorganization that can take place with, or without, an increase in brain size and the potential for evolving internal functional interactions, loops or modules (Oxnard 2004) becomes realized. Primate cortices may, indeed, represent “veritable hotbed(s) of evolutionary reorganization” (Preuss 2001 p 140). As students of paleoneurology have discerned, however, the high intelligence of today’s primates flowered from trends in primate brain evolution.
that reach back into deep time (Radinsky 1974). Given the complexities involved in disentangling the evolutionary dynamics of increasing brain size from the intricate (and often hidden) subtleties of neurological reorganization, that insight is somehow very satisfying.

References


10 Primate Life Histories

Elke Zimmermann · Ute Radespiel

Abstract
The life history of any species is determined by traits that characterize its developmental and reproductive rate as well as the reproductive effort spent over lifetime. In this chapter, we will present an overview of our current knowledge on the diversity of primate life history. We will explore potential links between life history and major biological factors, which are suggested as a partial explanation for the existing interspecific variations in life history. Furthermore, we will outline general principles and current hypotheses on the evolution of primate life history. Our review will show that extant primates ranging from nocturnal ancestral primates to apes provide an important biological substrate to illuminate evolutionary roots and selective forces that shaped our own life history.

10.1 Introduction
The life history of any species is determined by traits that characterize its developmental and reproductive rates as well as the reproductive effort spent over its lifetime (Stearns 1992). For mammals, these traits are usually expressed in the gestation length, number and size of offspring at birth, body mass and age at weaning, patterns of postnatal growth, age at first reproduction, interbirth interval, and life span. Life history pattern and variation can be studied on an interspecific or intraspecific level.

Intraspecific variations arise since life history traits have been shown to be modified by ecology, and different populations of a single species can therefore be expected to show some differences in life history according to the specific ecological settings they are experiencing (Lee and Kappeler 2003). Life history traits of a given species should therefore be taken as predispositions toward certain ranges of potential values (Kappeler et al. 2003), and these ranges can, on the other hand, be taken to explore the modulating effects of different socio-ecological parameters on life history. Broad interspecific comparisons of life history traits, on the other hand, form the basis for most existing life history models. Both levels of comparisons will be employed in this chapter in order to illuminate the major evolutionary forces shaping primate life history pattern.
Many aspects of biological timing show a systematic covariation with body mass (Calder 1984), such that body mass has been recognized to be a major predictor of life history variation among species. For example, a large species usually takes longer to grow to maturity and will have larger neonates than does a small one (Harvey et al. 1987; Ross 1998). The relationship with body mass is not an isometric function in which life history traits vary in direct proportion to body mass. Instead, a trait \( P \) will typically vary in an allometric fashion \( W \) that is based on the formula: \( P = aW^b \). The constant \( a \) (allometric coefficient) and \( b \) (allometric exponent) describe the specific nature of the scaling relationship in any given case. If this formula is logarithmically transformed, it describes a straight line with a slope \( b \). Due to fundamental interdependencies among body mass and most biological traits, allometric analyses are obligatory when exploring variations in primate life histories and are taken into consideration in all parts of this chapter. In allometric analyses, it is commonly observed that the data are divided into two or more subsets that show a similar scaling trend (allometric exponent or slope value) but are vertically separated (i.e., show different intercepts). These subsets can be referred to as “grades” and the vertical separation between them can be termed a “grade shift” (Purvis et al. 2003).

In this chapter, we will present an overview of our current knowledge on life history diversity among extant primates (including man), explore potential links between life history and major biological factors, which are suggested as a partial explanation for the existing interspecific variations, and outline general principles and current hypotheses of life history evolution.

10.2 General features of primate life histories in comparison to other mammals

The “speed” of life, or reproductive turnover, is a central concept in the field of life history research (Stearns 1992). Primate life histories lie at the slow end of the fast-slow continuum that has been described for mammals in general (Harvey and Clutton-Brock 1985; Read and Harvey 1989; Charnov and Berrigan 1993; Ross 1998). This is expressed in altered allometric relationships for many primate life history traits in comparison to those of other mammals. In particular, primates have longer gestation periods, smaller litter sizes, larger neonates, slower postnatal growth rates, a later age at first reproduction, and a longer life span than do most other mammals of the same body weight (Figure 10.1; Martin and MacLarnon 1985, 1988; Charnov 1991; Lee et al. 1991; Charnov and Berrigan 1993; Ross 1998). As a consequence, primates possess an extended period of
infancy and juvenility and have lower reproductive rates in comparison to other mammals of the same size (Charnov 1993; Charnov and Berrigan 1993).

Many hypotheses have been suggested that may explain why primates are so different. They can be broadly categorized into three groups of related arguments.

The first group of explanations deals with the evolutionary relationship among life history variables and brain size. Allometric analyses have shown that large-brained primates generally have long gestation periods, slow, prolonged growth periods, late sexual maturation, and long lives (Harvey et al. 1987; Allman et al. 1993; Charnov and Berrigan 1993; Hakeem et al. 1996; Barton 1999; Ross and Jones 1999; Ross 2003). Existing brain size hypotheses aim to explain the underlying evolutionary pathways (reviewed in Deaner et al. 2003, see below).

The second group of explanations is related to different mortality schedules that are environmentally imposed (Promislow and Harvey 1990; Charnov 1991, 1993; Stearns 1992; Janson and van Schaik 1993; Ross and Jones 1999). Primate-specific avenues have been suggested to act in the form of either a relatively high juvenile mortality that favors the evolution of slow growth, i.e., extended periods of juvenility (Janson and van Schaik 1993), or relatively low rates of adult mortality that may be connected to a late age of first reproduction and therefore delayed maturation (Promislow and Harvey 1990; Stearns 1992; Charnov 1993, see below). The third group of explanatory hypotheses stress the importance of ecological factors (i.e., diet, predation) on productive rates (birth rates, age at

Figure 10.1
Relationship between body mass and the two life history parameters life span (LS) and age at sexual maturity (SM) for primates and other mammals (modified after Charnov and Berrigan 1993)
first reproduction) and therefore the speed of life (Rowell and Richards 1979; Eisenberg 1981; Ross 1988, 1992a, b; van Schaik and Deaner 2002, see below). It should be noted that the second and the third groups of hypotheses may be partly connected via the parameter of mortality.

### 10.3 Variability in primate life history traits

The previous considerations contrasted primate life histories as a whole with those of other mammals. This generalization disregards the finding that primates themselves contain a broad variety of life history pattern (Figure 10.2).

- **Figure 10.2**
  Diversity of primate life histories (LS = life span, ARP = adult reproductive period, J = juvenility, L = lactation, GL = gestation length (modified after Schultz 1969))

#### 10.3.1 Gestation length

Gestation length in primates (see Appendix) varies considerably from 57 days in the brown mouse lemur (*Microcebus rufus*) to 285 days in the gorilla (*Gorilla*...
Gestation length is correlated with adult female body mass \( (r = 0.74) \) and even more so with neonatal body mass \( (r = 0.82) \) (Harvey et al. 1987). Gestation length in the order Primates seems to follow quite well the allometric equation \( G = k_g W_m^{0.10} \) (Martin and MacLarnon 1990) with \( W_m \) indicating maternal body mass. Correlations between the developmental state of the young and life history variables indicate that species with poorly developed (relatively altricial) young characteristically have small body mass and relatively short gestation periods, have limited motor and thermoregulatory abilities at birth, are nocturnal, and use nests (Martin 1990; Ross 2003). Species with more precocial young tend to have the opposite set of characteristics.

### 10.3.2 Neonatal body mass

Neonatal body mass ranges from about 4 g in *Microcebus rufus* to more than 3,000 g in *Homo sapiens* (see Appendix) and is highly correlated with adult body mass \( (r = 0.97) \) and neonatal brain mass \( (r = 0.99) \), i.e., large mothers produce large neonates with large brains (Harvey et al. 1987). Despite this high correlation, there is additional variation in relative neonatal body mass that requires an explanation. Besides the obvious connection to the altricial–precocial dichotomy, one parameter has repeatedly received attention in this context and this is litter size. It had been hypothesized that litter size should correlate negatively with neonatal body mass (Leutenegger 1973). Although this has been disputed by some authors (Harvey et al. 1987), others have found statistical support for this relationship, at least at an intraspecific level (Ross and Jones 1999). In general, however, it must be stated that the reduction in litter size and the “decision” for few and large infants was probably made very early in primate evolution, which limits its observable variation in the primate order. Other arguments have emphasized that the species-specific mode of placentation may constrain neonatal mass (Leutenegger 1973, 1976). Species with hemochorial placentas (anthropoid primates) may be able to nourish their fetus longer (and therefore to larger size) than species with an epitheliochorial placenta (most lemurs and lorises). Although this may explain some neonatal mass variations in primates, this does not hold for other mammalian orders (Martin 1984) and its general explanatory value is therefore unclear.

### 10.3.3 Litter size

The litter size in primate species is generally small and ranges from one to a maximum of four (average: 1–2, see Appendix). A litter size of one is typical for
most primates, and a litter size of two is the rule only for several mostly nocturnal strepsirhines and most callitrichids. A phylogenetic reconstruction revealed that the ancestral primate most likely had only a litter size of one and that litter size increased again secondarily several times (Kappeler 1998; see also Leutenegger 1979). Two different avenues have been suggested for the evolution of larger litters within the primate lineage. Larger litters occur either together with relatively altricial development when infants still have their eyes closed at birth and do not cling to their mother’s fur continuously (parkers in the strepsirhines, see mode of infant care). The alternative route has evolved in the family Callitrichidae, where relatively precocial twin litters are carried and raised not only by their mothers but with high proportions of allocare (Goldizen 1987; Ross 1991, 2003; Rothe et al. 1993; Dunbar 1995; Garber 1997; Bales et al. 2000; but see Mitani and Watts 1997). In view of the high energetic demands of twin production coupled with a postpartum oestrous in callitrichids, Dunbar (1995) and Ross (1991) suggested that the evolution of paternal care in this group preceded the evolution of twinning. An alternative explanation was offered by Chapman et al. (1990) who found a positive relationship between litter size and the proportion of insects in primate diets in a comparative study that included 70 primate species. They argued that insects may provide crucial energy supply in times of seasonal food shortage, which may then allow certain lineages to produce larger litters.

10.3.4 Age and body mass at weaning

Weaning is not a simple event but is instead a process that may occur over an extended period of time. Definitions of weaning may be as different as the “first intake of solid food,” the “beginning of maternal control of suckling,” the “complete termination of suckling,” or the “mother’s resumption of sexual activity” (reviewed in Lee 1996). As a general approximation, weaning age in primates ranges from about 40 days in *Microcebus murinus* to about 1,702 days (4.7 years) in *Pan troglodytes* (see Appendix). The size at weaning ranges from about 33 g in *Microcebus murinus* to about 11 kg in *Pongo pygmaeus*. The age of weaning is highly correlated with neonatal body mass ($r = 0.94$) and with adult female body mass ($r = 0.91$), i.e., large mothers produce large neonates that are weaned later than those of smaller mothers (Harvey et al. 1987). However, when the effect of the mother’s body mass is removed, only a relatively small proportion of the variance in weaning age can be explained by neonatal mass (Harvey et al. 1987; Lee et al. 1991). Weaning age, however, scales positively with weaning mass.
which is again highly proportional to neonatal mass (Lee et al. 1991). This relationship appears to be relatively constant, such that when an infant reaches about four times its birth weight, it is weaned irrespective of the duration of lactation (Lee et al. 1991). Another general pattern has been suggested with the mass at weaning equaling about one-third of the adult body mass (Lee et al. 1991; Lee 1996), although this relationship is debated (Godfrey et al. 2003; Purvis et al. 2003). It is apparent from many allometric analyses that primates show a considerable variation in the duration of lactation (infants can be weaned at an earlier or later age) but not in the body mass that is achieved at weaning. Lactation itself is a costly process for the mother who has to convert maternal energy to milk. It requires, for example, additional energetic intake at a rate of approximately 1.3 times the normal intake for humans or 1.5 times for baboons (Prentice and Whitehead 1987; Altmann and Samuels 1992). Mothers should balance these costs over time, and trade-offs should take into account not only the survival probabilities of the current offspring of a given mass and the energetic demands of mother and offspring during lactation but also survival costs for the mother and the relationship between weaning age, subsequent interbirth interval, and therefore future reproductive rates (Lee 1999; Ross and MacLarnon 2000). Factors that have been suggested to shape the species-specific evolutionary pathways in postnatal growth rates and weaning age are brain growth pattern as a metabolic constraint on infant somatic growth (Martin 1996; Lee 1999), environmental variability acting via differential juvenile mortality on maternal investment (age at weaning) and growth rates to weaning (Ross 1988; Leigh 1994; Garber and Leigh 1997; Lee 1999), and the extent of available allocare (Ross and MacLarnon 2000; Ross 2003).

### 10.3.5 Pattern of postnatal growth

Compared to other mammals, primates have the longest juvenile periods for their body size (Pereira and Fairbanks 1993). They range from 175 days in Galago moholi to 12.5 years in Homo sapiens (see Appendix). Juvenility that spans the time from weaning to sexual maturity, however, is not necessarily a period of continuous growth and acquisition of skills. Two major ways have been identified in which growth may vary within and among primate species (Leigh 1994, 1996): Postnatal body mass growth rates may slowly decrease from birth to sexual maturity, as can be seen, for example, in the common marmoset (Callithrix jacchus, Figure 10.3) or in females of some larger species such as the hamadryas
baboon (Papio hamadryas). The second pattern consists of slow growth rates early in life followed by sharp accelerations, so-called growth spurts, prior to sexual maturation. This pattern is characteristic for sexually dimorphic primate species such as baboons (Leigh 1996). Combinations of both basic patterns have generated a great variety of different growth-rate curves in primates. However, they do not correlate well either with other life history variables or with body mass, brain growth rates or with dental development (Pereira and Leigh 2003). For example, male Cercopithecus diana attains relatively small body mass after growing longer than male Colobus guereza (Pereira and Leigh 2003). It has been stated that such ontogenetic divergencies presumably reflect life history adaptations to contrasting socioecologies (Leigh and Shea 1995; Pereira 1995; Pereira and Leigh 2003). Among the parameters that are discussed as shaping growth curves are risks imposed on juveniles by predation or intraspecific competition for food, dominance, and social partners (Janson and van Schaik 1993; Leigh 1994; Godfrey et al. 2003; Pereira and Leigh 2003). Intrasexual competition has been shown, for example, to accelerate or extend male growth (bimaturism: Jarman 1983) and can even lead to alternative routes of reproduction (Utami et al. 2002; Wich et al. 2004 for the Sumatran orangutan, Pongo abelii).

10.3.6 Age at first reproduction

The age at first reproduction varies greatly among primates from about 8 months for Galago moholi or Microcebus murinus to more than 14 years for Pongo
Pygmaeus (see Appendix). The age at first reproduction marks a very important step in life since it affects the length of time available for reproduction and therefore limits the reproductive lifespan of males and females. Furthermore, it has an enormous influence on the intrinsic rate of population increase. In comparison to other mammals, maturity is reached later in primates (Ross 1988; Charnov and Berrigan 1993) and several hypotheses have been developed to explain this extended juvenility. First, allometric influences of adult body size and adult brain size must be addressed. Although there is a high correlation between adult body size and age at first reproduction (Pagel and Harvey 1993), this seems to be largely based on the effect of adult brain size (Harvey et al. 1987; Ross 2003). If this effect is removed, adult body mass and age at first reproduction are no longer significantly correlated. Large brains have been confirmed independently several times to correlate with late maturation (Harvey et al. 1987; Allman and Hasenstaub 1999; Barton 1999; Ross and Jones 1999; Ross 2003). The numerous correlations among the different life history variables mean that large-brained primates generally have slow, prolonged growth periods and late sexual maturation (Harvey et al. 1987; Charnov and Berrigan 1993). Deaner et al. (2003) have recently critically reviewed the available hypotheses that may account for this relationship. Other available explanations stress the importance of extended juvenility for the acquisition of learning skills or social maturation that may be enhanced in the primate lineage but may still be differentially expressed among different species (Harvey et al. 1987; Pagel and Harvey 1993). It is, however, unlikely that those parameters evolved fully independently of brain size. Finally, there has been some support for the hypothesis that juvenile mortality rates are correlated negatively with the length of the juvenile period, i.e., primates with high rates of prereproductive mortality reproduce at an early age and thereby speed up reproduction (Ross and Jones 1999).

### 10.3.7 Interbirth interval

The interbirth intervals of nonhuman primates range from 3 months in Microcebus murinus and Microcebus rufus up to 72 months in Pongo pygmaeus (see Appendix). The interbirth interval, and its variance, is strongly influenced by maternal mass (Harvey et al. 1987; Lee 1999), i.e., large mothers have longer interbirth intervals and have a greater potential to extend it further than smaller mothers. The minimum interbirth interval is usually given by the joined duration of gestation and lactation since in many species lactation prevents ovulation
(lactation amenorrhea). This phenomenon is adaptive if lactation costs are too high to be combined with the costs of gestation or if lactation is needed for a much longer time span than the gestation would take. In this case, females would risk having two generations of dependent offspring at the same time. Some primate species, such as some callitrichids (Dixson 1992; Tardif et al. 2003) or mouse lemurs (Schmelting et al. 2000), have evolved a postpartum oestrus. In these species, infant development is relatively fast and females can afford to lactate and resume reproduction simultaneously due to an early age of weaning.

Two other parameters have been identified that may shape the length of interbirth intervals. A very important one is seasonality in food availability that may constrain reproductive rates and even force species into a yearly cycle of reproduction (Lindburg 1987; Di Bitetti and Janson 2000). For example, gray mouse lemurs (*Microcebus murinus*) have been shown to reproduce twice per season in an area of 1,500-mm rain per year (Schmelting et al. 2000), whereas they produce only one litter in an area of about 800-mm rain per year (Eberle and Kappeler 2004). Within social groups of the same species, interbirth intervals also depend on maternal condition. It could be shown, for example, that better-fed (i.e., high-ranking) female olive baboons (*Papio anubis*) have shorter interbirth intervals and higher infant survival than low-ranking females (Packer et al. 1995). Similar results have been found in other species (Cheney et al. 1988 for *Cercopithecus aethiops*; Lee and Bowman 1995 for *Macaca fuscata*).

### 10.3.8 Life span

Primate life span is notoriously difficult to measure, since either the age of captive animals is often well known but may not necessarily reflect the life span that is typically achieved under natural conditions, and/or its determination in the field requires a long-term commitment of researchers to individually known populations that may span even more than a whole scientific career. Available life span data of nonhuman primates range from about 9 years in fat-tailed lemurs (*Cheirogaleus* spp.), Calabar angwantibos (*Arctocebus calabarensis*), or the Goeldi’s monkey (*Callimico goeldii*) up to about 50 years in orangutans (*Pongo pygmaeus*) (see Appendix). Thus, in comparison to other mammals, primates have longer lives for their size (Charnov and Berrigan 1993; Jones and MacLarnon 2001). Allometric analyses showed that life span within the primate lineage is positively correlated with body mass and brain size (Harvey et al. 1987; Austad and Fischer 1992; Allman et al. 1993; Allman 1995; Deaner et al. 2003).
Brain size seems to be more influential than body size since the partial correlation between life span and brain size remains positive after the effects of body size have been removed but not vice versa (Sacher 1959; Harvey et al. 1987; Allman et al. 1993; Deaner et al. 2003). In addition, age at maturity correlates strongly with life span, and it has even been suggested that life span and brain size may be primarily linked to age at maturity and may be only secondarily correlated with each other via the intervening variable age at maturity (Harvey et al. 1987). Both, life span and age at maturity affect and limit the time period available for reproduction, i.e., the reproductive life span. Reproductive life spans can be quite similar even among species with a divergent pattern of postnatal growth and maturation if, for example, longer lives may compensate for late maturation or vice versa.

The general question is, why did longevity evolve in primates and how can interspecific differences in primate life span be explained? In other words, what are the selective benefits of a long old age? In evolutionary terms, benefits are usually expressed as fitness consequences. On an individual level, an extended life span may result in a larger number of offspring and may therefore provide direct fitness benefits (Fedigan and Pavelka 2001 for *Macaca fuscata*). On the species level, however, a longer life span does not necessarily relate to a higher number of surviving offspring. A female chimpanzee may live up to 50 years but may have only about seven offspring due to late maturation and long interbirth intervals, whereas a rhesus macaque with a life span of 21 years or a mouse lemur of 10 years may each have given birth to about 18 offspring over their lifetime. If the number of offspring does not explain long life span, selective benefits are more likely to be found in offspring quality, which may be linked to other life history variables, in particular to an extended period of juvenility, delayed maturation, reduced rates of adult mortality, and larger brains.

Extrinsic mortality rates have repeatedly been investigated as a potential explanation for life history variations (see Age-specific Mortality Meder V3 Chap 12). Their connection to life span is suggestive since extrinsic mortality influences the probability of survival to old age and thereby determines the selection pressure upon genes that regulate somatic repair (Kirkwood and Rose 1991). This relation between mortality rates and longevity is supported by several theoretical and empirical studies (Promislow and Harvey 1990; Charnov 1991, 1993; Stearns 1992; Ricklefs 1998; Harvey and Purvis 1999; Ross and Jones 1999; Alvarez 2000; Blomquist et al. 2003).

An integrating explanation is offered by the “Cognitive Buffer Hypothesis” (Allman et al. 1993; Hakeem et al. 1996; Kaplan et al. 2000; reviewed in Deaner et al. 2003; see Brain Size).
10.4 Variables linked to primate life history traits

The evolution of primate life history is shaped by differences in brain sizes, phylogeny and by different ecological and social factors outlined in the following sections.

10.4.1 Brain size

Brain size and brain growth pattern are strongly interrelated with almost all life history traits. Large-brained primates usually have long gestation periods, high neonatal body mass, and give birth to few offspring that have a late age of first reproduction and long lives (Harvey et al. 1987; Austad and Fischer 1992; Allman et al. 1993; Charnov and Berrigan 1993; Purvis and Harvey 1995; Allman and Hasenstaub 1999). These relationships exist independently of body size (Sacher 1959; Allman et al. 1993; Martin 1996; Barton 1999;) and have been functionally linked either to high energetic costs of large brains (Foley and Lee 1992; Martin 1996; Ross and Jones 1999) or to long developmental processes (learning) during the juvenile period that require large brains (Harvey et al. 1987; Joffe 1997; Barton 1999; Dunbar 2003).

Deaner et al. (2003) recently presented a critical overview of the existing hypotheses that link brain size to primate life histories. Four hypotheses were judged to be compatible with the current state of knowledge. These are the “Maturational Constraint Hypothesis,” the “Cognitive Buffer Hypothesis,” the “Brain Malnutrition Risk Hypothesis,” and the “Delayed Benefits Hypothesis.” They are not mutually exclusive, but a final decision about their explanatory value cannot be made up to now.

The “Maturational Constraint Hypothesis” suggests that complex behaviors can be supported only by mature nervous systems and that an evolutionary increase in brain size often delays maturation (Altmann and Alberts 1987; Allman 1999; Allman and Hasenstaub 1999). It is based on the assumption that more complex patterns of behavior and neural connectivity are built on simpler ones and that complex behavior thus cannot emerge immediately, which is, for example, supported by some research on the evolution of neural networks (Quartz and Sejnowski 1977). Consequently, brain maturation may constrain behavioral and sexual maturation, and this assumption is supported by some neuroanatomical findings (Melnick and Pearl 1987; Nezu et al. 1997).

The “Cognitive Buffer Hypothesis” argues that large brains offer behavioral flexibility and higher learning capacities and thus serve to reduce extrinsic
mortality rates (e.g., deaths due to food shortages or predation). Large-brained animals are therefore better buffered against ecological danger and, if experiencing reduced mortality, possess a necessary predisposition for the selection of slower life histories. This hypothesis finds some support in the correlation between brain size and life span described by several authors (Harvey et al. 1987; Austad and Fischer 1992; Allman et al. 1993) and in a positive correlation of neocortex size (representing most structures involved in higher-order cognition) with life span (Deaner et al. 2003). Following this hypothesis, the main selective pressure was on brain size, and life history traits were either constrained by (e.g., extended juvenility or age at first reproduction) or coevolved together with brain size.

The “Brain Malnutrition Risk Hypothesis” concentrates on the high energetic costs of growing large brains (Martin 1981, 1996; Armstrong 1983; Leonard and Robertson 1992; Aiello and Wheeler 1995; Ross and Jones 1999; Kaplan et al. 2000) and is based on the assumptions (a) that the brain is more sensitive to energy shortage than the rest of the body and energy restrictions during brain development may lead to long-lasting or even permanent brain damage, which has to be prevented, and (b) that body growth trajectories lag behind brain growth trajectories in a general fashion that reflects a functional balance between behavioral capabilities and body size (Deaner et al. 2003). This model can explain the relationship between large brain size and late maturation, as the extension of the body growth period may allow the evolutionary increase in brain size. Furthermore, its predictions correspond well with the observed growth trajectories during lactation and after weaning including growth spurts that occur in certain species (Count 1947; Cheek 1975; Janson and van Schaik 1993; Leigh 1996; Bogin 1999; Deaner et al. 2003; Pereira and Leigh 2003).

The “Delayed Benefits Hypothesis” aims to explain the relationship between large brain size and longevity (Deaner et al. 2003). Large brains incur high costs (time and risks) but allow an animal to develop skills or knowledge that may lead to fitness benefits later in life. These benefits must exceed the costs of developing them, and long-lived animals have a longer time span to accumulate the benefits (Dukas 1998; Kaplan et al. 2000). Therefore, an evolutionary increase in longevity should increase the likelihood of an evolutionary increase in brain size. A strong correlation (without the outliers, bats) between brain size and life span across the eutherian mammalian orders supports the major prediction derived from this hypothesis (Deaner et al. 2003).

The four hypotheses differ not only in the life history trait which is most closely linked to brain size (age at first reproduction versus longevity) but also in their causal connection. For example, the “Maturational Constraint Hypothesis”
sees brain size as selective pressure for late maturation, whereas the “Brain Malnutrition Risk Hypothesis” argues that delayed maturation preceded the evolution of larger brains. Similarly, following the “Cognitive Buffer Hypothesis” larger brains facilitated the evolution of longevity, whereas in the “Delayed Benefits Hypothesis” longevity preceded the evolution of large brains. Future studies have to derive competing and testable predictions from these hypotheses to clarify the pathways that were relevant in primate evolution.

Recently, Leigh (2004) presented supportive evidence for an extension of the “Brain Malnutrition Risk Hypothesis,” since he identified two different maternal strategies within the primate lineage with regard to pre- and postnatal brain size development, metabolic needs, and age at maturation: the first strategy is followed by females who mature late with large body size and produce infants with relatively large brains (see also Harvey and Clutton-Brock 1985). Late maternal maturation was hypothesized to allow mothers to carry the metabolic costs of very early (fetal or infantile) brain growth through larger size. As a consequence, offspring may be more precocious in terms of cognition, locomotion, foraging, and social behavior. Squirrel monkeys or Old World monkeys as a group appear to follow this strategy (Leigh 2004). The second strategy consists of early maturation, coupled with a slow postnatal brain growth pattern over a long period of time, which may free the mother from extensive postnatal investment by “distributing” the costs of offspring brain growth to the offspring itself or to other group members. Tamarin females seem to follow this second strategy and give birth to offspring early, with short interbirth intervals (Leigh 2004). The differences between these two strategies would reflect different metabolic risks and trade-offs faced by various species at different stages of their life cycles.

10.4.2 Phylogeny

Phylogeny has long been recognized to constrain the evolution of life history. Thus closely related taxa, such as congeneric species, may show similar life history traits because they are likely to differ less in body size and ecology than more distantly related species (Harvey and Pagel 1991; but see Kappeler 1996). Primate life histories are highly conservative within lineages and show two major clades where strepsirhines as a group lie more to the fast end of the fast-slow continuum than haplorhines (Martin and MacLarnon 1990; Martin 2003; Purvis et al. 2003). Species within each clade share similar life history traits. Neonatal body mass of strepsirhines, for example, is almost three times below that for haplorhines of the same body size (Leutenegger 1973; Harvey et al. 1987; Martin 2003),
but strepsirhines tend to have more infants per litter (Chapman et al. 1990). Strepsirhine mothers allocate less energy to prenatal growth per unit time than haplorhines (Martin and MacLarnon 1988; Ross 1988; Young et al. 1990), whereas the two major haplorhine groups, the New World and the Old World monkeys, do not differ in this respect from each other (Ross 1991). After birth, the situation is reversed for the two clades, i.e., strepsirhines grow faster than haplorhines (Martin 2003; Figure 10.4). Differences in reproductive output of females between the two clades have led to the suggestion that haplorhines trade time for energetic investment in infant growth while strepsirhines opt for numbers (Lee and Kappeler 2003). Furthermore, relative brain size in strepsirhines is lower than in haplorhines (Martin 1990; Barton 1999; Deaner et al. 2003). Recently, Purvis et al. (2003) provided evidence for a mosaic evolution of life history patterns across five major primate clades (strepsirhines, platyrrhines, cercopithecines, colobines, and hominoids). The rate of accumulation of life history variation differed among the clades in a pattern that seems to correlate with their ecological diversification, i.e., strepsirhines showing the highest rate. Different phylogenetic groups have different patterns of variations, suggesting that they are moving differently on the fast-slow continuum, i.e., some by changing litter sizes, others by varying the age of sexual maturity, and so on. They found, however, no evidence for the claim that life history shaped primate phylogeny by affecting speciation or extinction rates.

Various methods have been introduced to remove the so-called confounding effect of phylogeny in comparative studies (Harvey and Pagel 1991; Purvis et al. 2003)

Figure 10.4
Grades of primate litter growth rates (modified after Ross 1998)
so that a potential covariation among life history traits and their evolutionary interplay with demographic, ecological, and social factors may be assessed.

### 10.4.3 Age-specific mortality rates

Extrinsic mortality rates during different phases of the life cycle have been used repeatedly to explain life history variations. Findings of Promislow and Harvey (1990) suggested, for example, that in mammals both high juvenile mortality and high adult mortality are linked to an early age at first reproduction when the effects of body size are removed. Charnov (1993) developed a comprehensive and most influential model that explains life history variation and in particular age at first reproduction as a consequence of adult mortality rates, which are in turn determined by the environment. When adult mortality is high, animals should mature early in order to minimize the risk of dying before reproducing. The Charnov-model is based on a growth law that constrains the relationship between age and body mass before maturity and is able to explain the allometric relationships of many life history traits with body mass, but recent studies shed doubts on its general explanatory value for life history variations in primates (see Purvis et al. 2003 for discussion). Despite this criticism, the general influence of mortality rates on life history pattern and on life span and age at first reproduction in particular is undisputed and is meanwhile supported by many studies (Janson and van Schaik 1993; Owens and Bennett 1995; Purvis and Harvey 1995; Charnov 1997; Kozlowski and Weiner 1997; Ricklefs 1998; Harvey and Purvis 1999; Ross and Jones 1999; Alvarez 2000) even if the exact mathematical model for their relationship is not yet formulated. It can be expected that this relationship is a result of balancing the benefits of reproducing early with relatively high birth rates with the benefits of delayed maturity that leads to larger mothers that may be more successful in raising young, and the costs and survival risks that are connected with each of these strategies (Ross 1998; Alvarez 2000).

### 10.4.4 Diet

Primates feed on a variety of different food items. According to their specialization to fruits, leaves, and animal protein (e.g., insects, arachnids, small vertebrates, eggs), they are usually classified as frugivores, folivores, and faunivores (Fleagle 1999). Since variation in reproductive capacity of a species is strongly affected by the availability and quality of food (Ganzhorn et al. 2003), diet
has been hypothesized to influence life history variation between species in a predictable way. To date, however, results are inconclusive.

Diet appears to have little effect either on prenatal maternal investment (litter weight/gestation length) or on the intrinsic rate of natural increase of a population of anthropoids (Ross 1988). However, a clear association between body mass, ontogeny, and diet was found in another study (Leigh 1994). In the latter, infant growth rate was faster in folivores than in comparably sized non-folivores. In a recent study with a much larger sample size including strepsirhines and focusing on intra- and interspecific variations, no significant effect of diet on life history traits was revealed once body mass was removed (Lee and Kappeler 2003). It was shown, furthermore, that the duration of postnatal growth as reflected in the age of weaning was longer for frugivores among anthropoids, whereas it was longer for folivores among strepsirhines.

10.4.5 Predation

Predation is a major ecological factor influencing mortality patterns in primate populations. It is known to be dependent on body size, such that larger-bodied animals face a lower predation risk than smaller-bodied ones (Read and Harvey 1989; Isbell 1994; Janson 2003). It is therefore expected that predation varies according to body size and affects reproductive output in primate females.

In order to explore the effect of predation on life history, researchers have often used subjective assessments of predation risks (Janson 2003; Lee and Kappeler 2003). For example, predation risk was estimated based on the response to predators and on predator encounter rates and then ranked according to low, medium, and high (Lee and Kappeler 2003). However, only a clade difference on relative reproductive output has been found so far (Lee and Kappeler 2003). Thus, short gestation periods were associated with high risks of predation in strepsirhines. Neonatal body mass and mass at weaning varied as a function of predation risk in anthropoids, such that under high risk infants were born large but weaned small, minimizing not only the risks after birth but also maternal investment during the postnatal growth phase associated with high risks (Lee 1999). Recently, Janson (2003) introduced a new theoretical model linking life history variation between species with predation risk and social grouping pattern. He was puzzled by the observation that primate species with large body size or living in large cohesive bisexual groups both assumed to be adaptations to high predation risk tend to have low predation rates. He revealed that interspecies variations in longevity can explain these observations, assuming that predation
risk remains fairly constant across primates. If longevity increases with increasing body mass, as has been shown empirically, then larger species face the cost of predation over a longer time span than smaller ones and consequently will gain a larger fitness benefit by evolving sophisticated antipredator responses (e.g., increased sociality, alarm calls).

### 10.4.6 Social grouping pattern

Three major patterns of social organization are found in primates (Kappeler and van Schaik 2002). Neighborhood systems (Richard 1985; Radespiel 2000), in which individuals forage separately and either sleep alone or come together at the end of their activity period to form permanent and long-lasting unisex or mixed-sexed sleeping groups, are found within the nocturnal strepsirhines, in tarsiers as well as in one diurnal anthropoid, the orangutan. In these systems, home ranges may overlap within and between sexes. Dispersed and cohesive pairs (Müller and Thalmann 2000; Rasoloharijaona et al. 2003), where both sexes forage either alone or in synchronized association and mostly sleep together, are found in various strepsirhines, in tarsiids, in various New World monkeys, as well as in gibbons. Sexes share the same home range and defend it against conspecifics. Paternal care of both sexes is a widespread phenomenon within pair-living primates (see section 10.4.7). Group-living in cohesive bisexual groups with high variations in size, sex ratio, and temporal stability is the most common pattern of social organization in diurnal primates (Kappeler and van Schaik 2002). Variations in group size depend on the dispersal regime of a particular sex and species as well as on the benefits of foraging and sleeping together and the costs of feeding competition, predation, and infanticide in the respective environment.

Relative group size and residence strategy (both sexes disperse, male philopatry, female philopatry) is associated with relative reproductive output. In a recent study in which the effect of group size [categorized as small (<4), medium (4–20), and large (>20)] and residency on intra- and interspecific life history variations was examined (Lee and Kappeler 2003), large groups of group-living anthropoids had a smaller relative neonatal mass and a higher relative mass at weaning than the other groups. These differences in life history variation between groups of different sizes might be explained either by a differential mortality risk in infants or juveniles due to predation or infanticide (Janson and Goldsmith 1995; Hill and Lee 1998; Janson and van Schaik 2000) or by a different degree of within and between group-feeding competition (van Schaik and van Hooff 1983;
Furthermore, in group-living primates with male philopatry, a longer relative gestation period, a later age at first reproduction, and a longer birth interval was revealed (Lee and Kappeler 2003). It was speculated that lower rates of extrinsic mortality, suggested to be related to group-living, may facilitate a higher maternal investment.

10.4.7 Mode of infant care

Parental investment theory suggests that parental care should be allocated among offspring in ways that maximize the parent’s expected lifetime reproductive success (Clutton-Brock 1991). Since the energy that can be devoted to reproduction by an individual is limited, trade-offs between energy investments in infant growth reflected by different modes of infant care and birth rates were suggested to explain some variation in life histories. Primates show a diversity of infant care patterns, which are found to be linked to the reproductive output of mothers.

All nocturnal strepsirhines (except woolly lemurs) as well as tarsiers leave their infants for extended periods of time cached at safe shelters, e.g., tree holes, nests, or dense vegetation, during foraging (Kappeler 1998; Rasoloharijaona et al. 2000; Lutermann 2001; Ross 2003). When changing shelters, mothers carry their infants either orally or clinging to their belly or by both modes of transportation. Species displaying this mode of infant care are called parkers (Ross 2001). In contrast, two nocturnal genera of strepsirhines (*Loris, Avahi*), almost all cathemeral lemurs (except *Hapalemur*), all diurnal lemurs as well as all anthropoids (except *Procolobus verus*) maintain close body contact to their infants during foraging while transporting them clinging to their body (Nicolson 1987; Whitten 1987; Ross 2003). Species with this mode of infant care, with no observed nesting and parking, are called riders (Ross 2001). Anthropoid riders show various degrees of allocare where other females, juveniles, or adult males (often related to the infant) help in infant care.

Parkers with oral infant transport tend to have multiple offspring per litter with low litter mass, born after a short gestation time (Kappeler 1998; Ross 2001; Wrogemann et al. 2001; Ross 2003). In contrast, parkers who transport infants clinging to their belly have relative long gestation periods and a single offspring with relatively low body mass (Zimmermann 1989; Ross 2003), suggesting that nesting may be more closely associated to the birth of multiple offspring than to the small size of offspring (Ross 2003). Whether a short gestation period selects for using nests or tree holes to rear offspring or vice versa is not yet clear.
Riding evolved repeatedly in different lineages of primates (Kappeler 1998; Ross 2003) and was most likely selected for reducing mortality risks in infants. Once evolved, it has been conserved, perhaps due to physiological coadaptations. For example, riders have developed a less energy rich milk compared to parkers (Tilden and Oftedal 1997). Riders are usually larger than related parkers suggesting that the evolution of large body size and infant fur-clinging may be linked (Kappeler 1998; Ross 2001). Riders have the same birth rates as parkers but wean and breed later when adjusted to body size and phylogeny. They also maintain smaller home ranges than parkers of the same mass (Ross 2001), indicating that carrying infants during foraging incurs energetic and reproductive costs.

Allocare is described in a variety of anthropoid riders (Nicolson 1987; Whitten 1987; Ross 2003) and has been linked to female reproductive rates. It gives mothers the possibility to enhance their reproductive potential (Mitani and Watts 1997; Ross and MacLarnon 2000). Species with high amounts of allocare show more rapid postnatal growth and wean their infants at a younger age (but at the same relative size) than those with low allocare (of the same body mass, Figure 10.5). This early weaning allows higher birth rates but does not lead to earlier matura-

10.5 Human life history

Humans (Homo sapiens) differ fundamentally from all other primates in four life history traits: neonatal weight, age at first reproduction, interbirth interval,
and life span (Smith and Tompkins 1995; Kaplan 2002; Hawkes et al. 2003). Compared to apes, human babies are relatively large at birth, but they grow more slowly and have a fairly extended childhood (period following infancy when youngster is weaned but still depends on caregivers for feeding and protection [Bogin 1997]). Furthermore, humans reach reproductive age much later, have shorter interbirth intervals and a higher longevity than apes (Leigh and Park 1998; Kaplan et al. 2000; Mace 2000).

Maximum life span in humans is reported to range between 90 and 122 years (Weiss 1981), which is about twice as long as in chimpanzees and gorillas (Kaplan and Robson 2002). A large proportion of this time is postreproductive in women, i.e., after woman have entered menopause, implying that total life span and fertile life span are disconnected in humans. For example, an 85-year-old woman has spent about 41% of her life in a postreproductive state, which is very much in contrast to nonhuman primates, whose postreproductive periods are usually short and coincide with frailty and somatic deterioration (Hawkes et al. 1998; Johnson and Kapsalis 1998; Pavelka and Fedigan 1999). Whereas the fertile life span of humans is quite similar to that of great apes (Alvarez 2000), woman can have a much higher fecundity due to much shorter interbirth intervals (Hawkes et al. 2003).

In addition to these specialized life history patterns, human brain size is three- to fourfold larger than in chimpanzees and gorillas (Kaplan and Robson 2002) and much higher than expected for primates when adjusted to body size (Pagel and Harvey 1989). The human brain grows faster and continues to grow longer after birth (Bogin 1999; Kaplan 2002).

Furthermore, humans show broad intraspecies variations in growth rate, age of maturation, birth rate, age-specific mortality, and senescence, which not only depends on ecological factors but also on food technology, medical care, and culture (Mace 2000; Kaplan 2002; Hawkes et al. 2003). In developed countries during the last 150 years, the average age at menarche among girls has decreased by about 4 years; birth rate in women has changed from six to about two children per family, and the percentage of people reaching more than 100 years of age has increased manyfold (Kaplan 2002).

To date, there has been much lively debate, but no consensus, on how to explain these human peculiarities in life history patterns. Two recently developed models provide two alternative explanations for selection pressures shaping the evolution of the exceptionally long life span in humans with the emergence of menopause in women.

One model, the “embodied capital hypothesis” (Kaplan et al. 2000, 2003; Kaplan and Robson 2002), proposes that the prolonged longevity and the large brain size of humans are products of coevolutionary selection favored by the
exploration of a new, cognitively challenging feeding niche. Ecological factors increasing the productivity of investments in developmental embodied capital, such as the brain, should increase both time spent as a child and investments to reduce mortality.

Supporting evidence for this hypothesis was gained by studies on foraging activity, net food production, and mortality in chimpanzee and contemporary hunter-gatherer populations (Kaplan et al. 2000; Kaplan and Robson 2002). Hunter-gatherers showed a much lower mortality rate during child- and adulthood than chimpanzees along with a quite different net food production profile (difference between food produced and food consumed; see Figure 10.6). Chimpanzee and human infants depended nutritionally on their mothers until weaning displaying a negative net food production. In contrast to chimpanzees, humans showed an extended juvenility with further nutritional provisioning by parents up to puberty lowering juvenile mortality. This extraordinary long juvenile dependency paid off later in life when the highly efficient foraging strategies of adults allowed them to obtain a much greater quantity of valuable food than required for personal subsistence, which is reflected in a positive net food production. Whereas these results suggest why longevity may have been extended and brain size enlarged in humans compared to apes, the recent discovery of the apolipoprotein E Allel-system, which affects neurite growth and aging, offers a potential candidate for the genetic basis of such a coevolutionary process (Kaplan and Robson 2002; Finch and Stanford 2004).

Figure 10.6
Net food production and mortality in human hunter-gatherer populations and chimpanzee populations (modified after Kaplan and Robson 2002)
The other model, the “grandmother hypothesis” (Hawkes et al. 1998, 2003; Alvarez 2002; Lahdenperä et al. 2004), is based on Charnov (1993) and assumes that prolonged postreproductive life span will evolve when postreproductive females gain greater fitness by increasing the success of their offspring than by continuing to breed themselves. The greater longevity favored delayed maturity because lower adult mortality rates reduce the fitness costs of prolonged growth.

Empirical support for this hypothesis comes from some contemporary and historical populations (Blurton Jones et al. 1989; Mace 2000; Jamison et al. 2002; Voland and Beise 2002; Hawkes 2003), but not all (Hill and Hurtado 1991, 1996, 1999). These studies showed that grandparents may assist philopatric offspring by transferring knowledge and by participating in household tasks and child care. This help may increase offspring breeding probability, grandchildren nutrition and survival. A recent comprehensive study on fitness benefits of prolonged postreproductive life span in women using multigenerational individual-based data sets from historical farming communities in Finland \((n \sim 500 \text{ women})\) and Canada \((n \sim 2300 \text{ women})\) during the eighteenth and nineteenth century (Lahdenperä et al. 2004) provides additional support. Postreproductive women, living in the same household or close to their offspring, have more grandchildren, and greater fitness, because they enhance the lifetime reproductive success of their offspring by allowing them to breed earlier, more frequently and more successfully. Fitness benefits, however, disappear as the reproductive output of the offspring declines. Rates of female mortality accelerate as their offspring cease reproduction.

10.6 Conclusions

As shown in this chapter, primate life histories are among the slowest in mammals and nevertheless show a fascinating diversity in pattern and variations within and between more than 250 extant species. The basic traits of life history, such as gestation length, neonatal body mass, number of offspring, age and body mass at weaning, postnatal growth, age at first reproduction, interbirth interval as well as life span, are linked in complex ways, and to varying degrees, across taxa by physiology (e.g., trade-offs in growth and reproduction) and genetics (e.g., pleiotropy). However, despite more than 30 years of intensive research, the causes of the unique slowness of primate life histories remain poorly understood. The three suggested major determinants, energetic costs of brain growth, age-specific mortality, and socioecology, explain in part the low developmental and reproductive rates of primates but do not yet deliver a comprehensive and conclusive
explanation. For example, we do not know whether the energetic costs for brain
growth or body growth drive life history evolution, or vice versa, or whether there
are unexplored factors driving their coevolution. Further progress in the devel-
opment of life history theory is necessary and further empirical data from
different primate lineages under different ecological settings, including primate
genomics, have to be compiled to convincingly identify targets of natural selec-
tion and to understand the full adaptive significance of life history variation
among primates.

Differences in basic life history traits between species are suggested to be
adaptive responses to past environments. As phylogenetic constraints they may
limit growth and mortality patterns of extant primate populations and as a
consequence affect their survival in rapidly changing environments. Most extant
nonhuman primate species are now at risk of extinction. A better knowledge of
their life history schedules and the limits of their phenotypic plasticity across
distribution ranges along with information on their population dynamics and
genetic variation may not only illuminate pathways for evolution but also help
in establishing effective conservation and management programs. Since respective
information on the link between distribution patterns, ecology, and life history is
available for only a few species, comparative long-term field studies applying similar
methodologies are urgently needed. These studies may also shed light on an
interesting paradox in life history research, namely that correlations for life history
traits sometimes diverge in intra- and interspecific comparisons (Martin 2003).

Finally, understanding the peculiarities of human evolution will require
further comparative research on nonhuman primates and their life histories.
Thus, for instance, recent socioecological studies have unraveled an unexpected
complexity in primate sociality, from solitary foraging and sleeping over dis-
persed or cohesive group living to more flexible fission–fusion societies. Since
these different degrees of sociality require quite different levels of brain complex-
ity for spatial and social memory, integrative and comparative studies on life
history, socioecology, and cognition within and across the different primate
lineages offer an excellent opportunity to illuminate our biological roots and
the selective forces that shaped our own life history.

Acknowledgments

We wish to thank Hartmut Rothe and Winfried Henke for the invitation to
contribute to this volume and all of the editors for helpful comments. Thanks
go furthermore to Elisabeth Engelke for editorial help and Rüdiger Brüning for
technical support in preparing the figures.
## Appendix Primate database for life history traits

<table>
<thead>
<tr>
<th>Family</th>
<th>Genus</th>
<th>Species</th>
<th>BM</th>
<th>AFR</th>
<th>GL</th>
<th>IBI</th>
<th>LS</th>
<th>NNM</th>
<th>WA</th>
<th>WM</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cheirogaleidae</td>
<td>Allocebus</td>
<td>trichotis</td>
<td>83</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cheirogaleidae</td>
<td>Cheirogaleus</td>
<td>major</td>
<td>356</td>
<td>71</td>
<td>2</td>
<td>2</td>
<td></td>
<td>18</td>
<td>70</td>
<td></td>
<td>15.0</td>
</tr>
<tr>
<td>Cheirogaleidae</td>
<td>Cheirogaleus</td>
<td>medius</td>
<td>139</td>
<td>1.19</td>
<td>62</td>
<td>12</td>
<td>2</td>
<td>12</td>
<td>61</td>
<td></td>
<td>9.0</td>
</tr>
<tr>
<td>Cheirogaleidae</td>
<td>Microcebus</td>
<td>bertha</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cheirogaleidae</td>
<td>Microcebus</td>
<td>griseorufus</td>
<td>63</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cheirogaleidae</td>
<td>Microcebus</td>
<td>murinus</td>
<td>63</td>
<td>0.67</td>
<td>60</td>
<td>37</td>
<td>2</td>
<td>5</td>
<td>40</td>
<td>332</td>
<td>15.5</td>
</tr>
<tr>
<td>Cheirogaleidae</td>
<td>Microcebus</td>
<td>myoxinus</td>
<td>49</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cheirogaleidae</td>
<td>Microcebus</td>
<td>ravelobensis</td>
<td>56</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cheirogaleidae</td>
<td>Microcebus</td>
<td>rufus</td>
<td>42</td>
<td>1.5</td>
<td>57</td>
<td>35</td>
<td>2</td>
<td>4</td>
<td>45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cheirogaleidae</td>
<td>Microcebus</td>
<td>sambiranensis</td>
<td>44</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cheirogaleidae</td>
<td>Microcebus</td>
<td>tavaratra</td>
<td>61</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cheirogaleidae</td>
<td>Mirza</td>
<td>coquereli</td>
<td>297</td>
<td>1</td>
<td>87</td>
<td>12</td>
<td>2</td>
<td>14</td>
<td></td>
<td></td>
<td>15.3</td>
</tr>
<tr>
<td>Cheirogaleidae</td>
<td>Phaner</td>
<td>furcifer</td>
<td>328</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12.0</td>
</tr>
<tr>
<td>Daubentoniidae</td>
<td>Daubentonia</td>
<td>madagascariensis</td>
<td>2,490</td>
<td>3.5</td>
<td>164</td>
<td>20</td>
<td>1</td>
<td>102</td>
<td>170</td>
<td>1,535</td>
<td>23.3</td>
</tr>
<tr>
<td>Lepilemuridae</td>
<td>Lepilemur</td>
<td>dorsalis</td>
<td>2,572</td>
<td>169</td>
<td>147</td>
<td>365</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lepilemuridae</td>
<td>Lepilemur</td>
<td>edwardsi</td>
<td>934</td>
<td>954</td>
<td>1</td>
<td>27</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lepilemuridae</td>
<td>Lepilemur</td>
<td>leucopus</td>
<td>594</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lepilemuridae</td>
<td>Lepilemur</td>
<td>mustelinus</td>
<td>895</td>
<td>895</td>
<td>1</td>
<td>27</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lepilemuridae</td>
<td>Lepilemur</td>
<td>ruficaudatus</td>
<td>779</td>
<td>1.63</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>120</td>
<td></td>
</tr>
<tr>
<td>Lepilemuridae</td>
<td>Lepilemur</td>
<td>septentrionalis</td>
<td>~750</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lemuridae</td>
<td>Eulemur</td>
<td>albocollaris</td>
<td>2,150</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lemuridae</td>
<td>Eulemur</td>
<td>coronatus</td>
<td>1,080</td>
<td>1,177</td>
<td>1</td>
<td>48</td>
<td></td>
<td>68</td>
<td>135</td>
<td>18.4</td>
<td></td>
</tr>
<tr>
<td>Lemuridae</td>
<td>Eulemur</td>
<td>fulvus</td>
<td>2,206</td>
<td>2,250</td>
<td>2.16</td>
<td>120</td>
<td>48</td>
<td>68</td>
<td>135</td>
<td>37.0</td>
<td></td>
</tr>
</tbody>
</table>

Notes: BM = body mass (g), AFR = adult female rank, GL = gestation length (days), IBI = interbirth interval (days), LS = litter size, NNM = number of nursing males, WA = weaning age (days), WM = weaning mass (g), L = longevity (years).
<table>
<thead>
<tr>
<th>Family</th>
<th>Genus</th>
<th>Species</th>
<th>BM</th>
<th>AFR</th>
<th>GL</th>
<th>IBI</th>
<th>LS</th>
<th>NNM</th>
<th>WA</th>
<th>WM</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lemuridae</td>
<td>Eulemur</td>
<td>macaco</td>
<td>1,760</td>
<td>1.66</td>
<td>129</td>
<td>12</td>
<td>1</td>
<td>50</td>
<td>135</td>
<td></td>
<td>27.1²</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1,793</td>
<td>2.18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Eulemur</td>
<td>mongoz</td>
<td>1,481</td>
<td>2.5</td>
<td>129</td>
<td>1</td>
<td>57</td>
<td>152</td>
<td></td>
<td></td>
<td>24.3¹³</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1,560</td>
<td>2.52</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Eulemur</td>
<td>rubriventer</td>
<td>1,940</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1,390</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hapalemur</td>
<td>griseus</td>
<td>670</td>
<td>2.38</td>
<td>140</td>
<td>11</td>
<td>1</td>
<td>~45</td>
<td>120</td>
<td></td>
<td>12.1²</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>790</td>
<td>2.92</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lemuridae</td>
<td>Hapalemur</td>
<td>griseus alaotrensis</td>
<td>~1,450</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hapalemur</td>
<td>simus</td>
<td>1,300</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lemur</td>
<td>catta</td>
<td>2,210</td>
<td>2.01</td>
<td>135</td>
<td>14</td>
<td>1</td>
<td>65</td>
<td>105</td>
<td></td>
<td>27.1²</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2,290</td>
<td>2.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lemuridae</td>
<td>Varecia</td>
<td>variegata</td>
<td>3,100</td>
<td>1.42</td>
<td>102</td>
<td>12</td>
<td>2</td>
<td>78</td>
<td>89</td>
<td>2,893</td>
<td>28.0¹³</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3,520</td>
<td>1.92</td>
<td>175</td>
<td></td>
<td>~87</td>
<td>90</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3,600</td>
<td>2.72</td>
<td></td>
<td></td>
<td>100</td>
<td>146</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indriidae</td>
<td>Avahi</td>
<td>laniger</td>
<td>875</td>
<td>2.58</td>
<td>12</td>
<td>1¹</td>
<td></td>
<td>150</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1,320</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indriidae</td>
<td>Avahi</td>
<td>occidentalis</td>
<td>777</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indriidae</td>
<td>Propithecus</td>
<td>tattersalli</td>
<td>3,500</td>
<td>4.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indriidae</td>
<td>Propithecus</td>
<td>verreauxi</td>
<td>2,950</td>
<td>3.5</td>
<td>140</td>
<td>12¹</td>
<td>1¹</td>
<td>103</td>
<td>180</td>
<td></td>
<td>18.2¹</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3,620</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indriidae</td>
<td>Propithecus</td>
<td>coquereli</td>
<td>3,244</td>
<td>4.2</td>
<td>141</td>
<td>12</td>
<td>1</td>
<td>103</td>
<td>182</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4,280</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indriidae</td>
<td>Propithecus</td>
<td>diadema</td>
<td>6,260</td>
<td>4</td>
<td>178</td>
<td>25</td>
<td>1¹</td>
<td>~135</td>
<td>183</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>672</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Indri</td>
<td>indri</td>
<td>6,240</td>
<td>159</td>
<td>30</td>
<td>1¹</td>
<td></td>
<td>363</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>6,840</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lorisidae</td>
<td>Arctocebus</td>
<td>aureus</td>
<td>~210</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Appendix** (continued)
<table>
<thead>
<tr>
<th>Family</th>
<th>Genus</th>
<th>Species</th>
<th>Births</th>
<th>Deaths</th>
<th>AvgSurv</th>
<th>Males</th>
<th>Females</th>
<th>Males</th>
<th>Females</th>
<th>Males</th>
<th>Females</th>
<th>1st Birth</th>
<th>1st Birth</th>
<th>1st Birth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lorisidae</td>
<td>Arctocebus</td>
<td>calabarensis</td>
<td>254</td>
<td>1</td>
<td>135</td>
<td>6</td>
<td>1</td>
<td>32</td>
<td>105</td>
<td>160</td>
<td>12.213</td>
<td>11</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>Lorisidae</td>
<td>Loris</td>
<td>lydekkerianus</td>
<td>269</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lorisidae</td>
<td>Loris</td>
<td>malabaricus</td>
<td>193</td>
<td>1</td>
<td>166</td>
<td>6</td>
<td>1</td>
<td>10</td>
<td>170</td>
<td>139</td>
<td>16.013</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lorisidae</td>
<td>Loris</td>
<td>tardigradus</td>
<td>255</td>
<td>1.5</td>
<td>166</td>
<td>6</td>
<td>1</td>
<td>10</td>
<td>170</td>
<td>139</td>
<td>16.013</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lorisidae</td>
<td>Nycticebus</td>
<td>bengalensis</td>
<td>95024</td>
<td>2.24</td>
<td>187</td>
<td>12</td>
<td>1</td>
<td>38</td>
<td>180</td>
<td>525</td>
<td>15.024</td>
<td>11</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>Lorisidae</td>
<td>Nycticebus</td>
<td>coucang</td>
<td>626</td>
<td>2.11</td>
<td>193</td>
<td>12</td>
<td>1</td>
<td>51</td>
<td>180</td>
<td>520</td>
<td>26.513</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lorisidae</td>
<td>Nycticebus</td>
<td>javanicus</td>
<td>798</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lorisidae</td>
<td>Nycticebus</td>
<td>menagensis</td>
<td>511</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lorisidae</td>
<td>Nycticebus</td>
<td>pygmaeus</td>
<td>~307</td>
<td>185</td>
<td>2</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lorisidae</td>
<td>Perodicticus</td>
<td>edwardsi</td>
<td>1,210</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lorisidae</td>
<td>Perodicticus</td>
<td>potto</td>
<td>836</td>
<td>2.03</td>
<td>195</td>
<td>12</td>
<td>1</td>
<td>33</td>
<td>150</td>
<td></td>
<td>26.013</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Galagidae</td>
<td>Euoticus</td>
<td>elegantulus</td>
<td>283</td>
<td>1352</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Galagidae</td>
<td>Euoticus</td>
<td>palidus</td>
<td>~300</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Galagidae</td>
<td>Galago</td>
<td>aleni</td>
<td>262</td>
<td>0.83</td>
<td>135</td>
<td>1217</td>
<td>1</td>
<td>24</td>
<td></td>
<td>8.01</td>
<td>10.024</td>
<td>24026</td>
<td>24026</td>
<td>24026</td>
</tr>
<tr>
<td>Galagidae</td>
<td>Galago</td>
<td>matschiei</td>
<td>212</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Galagidae</td>
<td>Galago</td>
<td>moholi</td>
<td>173</td>
<td>0.71</td>
<td>123</td>
<td>6</td>
<td>1</td>
<td>12</td>
<td>84</td>
<td>95</td>
<td>16.51</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Galagidae</td>
<td>Galago</td>
<td>senegalensis</td>
<td>195</td>
<td>1.4</td>
<td>142</td>
<td>6</td>
<td>1</td>
<td>19</td>
<td>98</td>
<td>150</td>
<td>16.02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Galagidae</td>
<td>Galago</td>
<td>demidoff</td>
<td>60</td>
<td>0.97</td>
<td>110</td>
<td>12</td>
<td>2</td>
<td>8</td>
<td></td>
<td>14.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Galagidae</td>
<td>Galagoides</td>
<td>thomasi</td>
<td>130</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Galagidae</td>
<td>Galagoides</td>
<td>zanzibaricus</td>
<td>132</td>
<td>1</td>
<td>126</td>
<td>6</td>
<td>1</td>
<td>~14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Galagidae</td>
<td>Otolemur</td>
<td>crassicaudatus</td>
<td>1,110</td>
<td>2.17</td>
<td>135</td>
<td>12</td>
<td>2</td>
<td>~43</td>
<td>135</td>
<td>500</td>
<td>18.813</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Galagidae</td>
<td>Otolemur</td>
<td>garnetti</td>
<td>721</td>
<td>1.58</td>
<td>132</td>
<td>7</td>
<td>1</td>
<td>49</td>
<td></td>
<td></td>
<td>15.013</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tarsiidae</td>
<td>Tarsius</td>
<td>bancanus</td>
<td>109</td>
<td>2.52</td>
<td>178</td>
<td>8</td>
<td>1</td>
<td>~24</td>
<td>79</td>
<td></td>
<td>12.01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tarsiidae</td>
<td>Tarsius</td>
<td>dianae</td>
<td>107</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tarsiidae</td>
<td>Tarsius</td>
<td>pumilis</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Primate life histories
## Appendix (continued)

<table>
<thead>
<tr>
<th>Family</th>
<th>Genus</th>
<th>Species</th>
<th>BM</th>
<th>AFR</th>
<th>GL</th>
<th>IBI</th>
<th>LS</th>
<th>NNM</th>
<th>WA</th>
<th>WM</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tarsiidae</td>
<td>Tarsius</td>
<td>spectrum</td>
<td>108</td>
<td>1.42</td>
<td>5</td>
<td>1</td>
<td>~23</td>
<td>68</td>
<td>12.0²</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tarsiidae</td>
<td>Tarsius</td>
<td>syrichta</td>
<td>120</td>
<td>180</td>
<td>1</td>
<td>26²</td>
<td>82</td>
<td>15.0¹³</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cebidae</td>
<td>Aotus</td>
<td>azarai</td>
<td>1,230</td>
<td>12</td>
<td>1</td>
<td></td>
<td></td>
<td>231</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cebidae</td>
<td>Aotus</td>
<td>lemurinus</td>
<td>874</td>
<td>133</td>
<td>12</td>
<td>1</td>
<td>98</td>
<td>75</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cebidae</td>
<td>Aotus</td>
<td>nancymae</td>
<td>780</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cebidae</td>
<td>Aotus</td>
<td>nigriceps</td>
<td>1,040</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cebidae</td>
<td>Aotus</td>
<td>trivirgatus</td>
<td>724</td>
<td>2.38</td>
<td>133</td>
<td>9¹</td>
<td>1¹</td>
<td>94</td>
<td>75</td>
<td>360</td>
<td>20.0¹</td>
</tr>
<tr>
<td>Cebidae</td>
<td>Aotus</td>
<td>vociferans</td>
<td>698</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>179</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cebidae</td>
<td>Cebus</td>
<td>albifrons</td>
<td>2,067</td>
<td>4.02</td>
<td>155</td>
<td>18</td>
<td>1¹</td>
<td>228</td>
<td>269</td>
<td>1,000³</td>
<td>44.0¹</td>
</tr>
<tr>
<td>Cebidae</td>
<td>Cebus</td>
<td>apella</td>
<td>2,201</td>
<td>5.5</td>
<td>154</td>
<td>22</td>
<td>1</td>
<td>197</td>
<td>261</td>
<td>1,000³</td>
<td>45.1¹³</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2,520</td>
<td>5.78</td>
<td></td>
<td></td>
<td></td>
<td>232</td>
<td>264</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cebidae</td>
<td>Cebus</td>
<td>capucinus</td>
<td>2,540</td>
<td>4</td>
<td>162</td>
<td>26</td>
<td>1</td>
<td>~230</td>
<td>510</td>
<td>1,350</td>
<td>54.8¹³</td>
</tr>
<tr>
<td>Cebidae</td>
<td>Cebus</td>
<td>olivaceus</td>
<td>2,520</td>
<td>6</td>
<td>26¹</td>
<td>1¹</td>
<td></td>
<td>730</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cebidae</td>
<td>Saimiri</td>
<td>boliviensis</td>
<td>711</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cebidae</td>
<td>Saimiri</td>
<td>oerstedii</td>
<td>680</td>
<td></td>
<td></td>
<td>1²</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cebidae</td>
<td>Saimiri</td>
<td>sciureus</td>
<td>662</td>
<td>2.5</td>
<td>170</td>
<td>9</td>
<td>1</td>
<td>106</td>
<td>168</td>
<td>418</td>
<td>27.0¹³</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>699</td>
<td>146</td>
<td>183</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>240</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cebidae</td>
<td>Saimiri</td>
<td>ustus</td>
<td>799</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cebidae</td>
<td>Saimiri</td>
<td>vanzolinii</td>
<td>650</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Callitrichidae</td>
<td>Callimico</td>
<td>goeldii</td>
<td>355</td>
<td>1.32</td>
<td>151</td>
<td>9</td>
<td>1</td>
<td>48</td>
<td>65</td>
<td>215</td>
<td>17.9¹³</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>~582</td>
<td>155</td>
<td></td>
<td></td>
<td>53</td>
<td>70</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Callitrichidae</td>
<td>Cebuella</td>
<td>pygmaea</td>
<td>79</td>
<td>1.88</td>
<td>137</td>
<td>6¹</td>
<td>2</td>
<td>12</td>
<td>90</td>
<td>70</td>
<td>18.0¹³</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>122</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Callitrichidae</td>
<td>Callithrix</td>
<td>aurita</td>
<td>429</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Callitrichidae</td>
<td>Callithrix</td>
<td>flaviceps</td>
<td>~406</td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Callitrichidae</td>
<td>Callithrix</td>
<td>geoffroyi</td>
<td>~359</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Family</td>
<td>Genus</td>
<td>Species</td>
<td>Code</td>
<td>Mean</td>
<td>SD</td>
<td>Min</td>
<td>Median</td>
<td>10</td>
<td>25</td>
<td>Median</td>
<td>75</td>
</tr>
<tr>
<td>------------</td>
<td>-----------</td>
<td>------------</td>
<td>------</td>
<td>-------</td>
<td>------</td>
<td>-----</td>
<td>--------</td>
<td>----</td>
<td>----</td>
<td>--------</td>
<td>----</td>
</tr>
</tbody>
</table>
| Callitrichidae | Callithrix | jacchus    | 287  | 1.37  | 148  | 6   | 2      | 27 | 60 | 128    | 16.7 | 13 
|             |           |            | 381  | 1.5   | 30   | 77  |        |     |     | 90     |    |     |     |
| Callitrichidae | Callithrix | kuhlii     | ~375 |       |      |     |        |     |     |        |    |     |     |
| Callitrichidae | Callithrix | penicillata| 307  |       |      |     |        |     |     | 2      |    |     |     |
| Callitrichidae | Leontopithecus | caissara | 572  |       |      |     |        |     |     |        |    |     |     |
| Callitrichidae | Leontopithecus | chrysomelas| 535  |       |      |     |        |     |     |        |    |     |     |
| Callitrichidae | Leontopithecus | rosalia    | 559  | 2.38  | 129  | 6   | 2      | 50 | 90 | 165    | 22.0 | 13 
|             |           |            | 598  | 2.42  | 62   |    |        |     |     |        |    |     |     |
| Callitrichidae | Saguinus   | bicolor    | 430  |       |      |     |        |     |     |        |    |     |     |
| Callitrichidae | Saguinus   | fuscicollis| 350  | 2.33  | 148  | 12  | 2      | 39 | 90 |        | 24.5 | 13 
| Callitrichidae | Saguinus   | geoffroyi  | 502  | 8     | 2    | 48  |        |     |     |        |    |     |     |
| Callitrichidae | Saguinus   | imperator  | 475  |       |      |     |        |     |     | 47     |    |     |     |
| Callitrichidae | Saguinus   | inustus    | 803  |       |      |     |        |     |     |        |    |     |     |
| Callitrichidae | Saguinus   | labiatus   | 520  |       | 145  | 10  | 2      | 38 |    |        |    |     |     |
| Callitrichidae | Saguinus   | leucopus   | 490  |       |      |     |        |     |     |        |    |     |     |
| Callitrichidae | Saguinus   | midas      | 558  | 2     | 7    | 2   | 40    | 70 | 13 |        | 13.0 | 2  
| Callitrichidae | Saguinus   | mystax     | 539  | 1.25  | 5    | 11  | 2      | 47 |    |        |    |     |     |
| Callitrichidae | Saguinus   | nigricolor | 350  | 2.33  | 1    | 44  | 77    | 175| 15.2|        |     |     |
| Callitrichidae | Saguinus   | oedipus    | 404  | 1.89  | 168  | 7   | 2      | 42 | 50 | 130    | 23.0 | 13 
| Callitrichidae | Mico       | argentatus | 353  | 1.67  | 7    | 2   | 36    | 120 | 130|        |     |     |
| Callitrichidae | Mico       | emiliae    | 330  |       |      |     |        |     |     |        |    |     |     |
| Callitrichidae | Mico       | humeralifer| 380  |       |      |     |        |     |     |        |    |     |     |
| Callitrichidae | Mico       | mauesi    | 398  |       |      |     |        |     |     |        |    |     |     |
| Callitrichidae | Mico       | nigriceps | 390  |       |      |     |        |     |     |        |    |     |     |
| Pitheciidae   | Cacajao    | calvus     | 2,880| 3.6   | 180  | 1   |        | 638|    |        | 22.3 | 13 
|               |            |            | 3,600|       |      |     |        |     |     |        |    |     |     |
| Pitheciidae   | Cacajao    | melanocephalus | 2,710 |   |     |     | |     |     |     |        |     |     |
| Pitheciidae   | Callicebus | brunneus   | 805  |       |      |     |        |     |     |        |    |     |     |
| Pitheciidae   | Callicebus | cupreus    | 1,120| 130   | 13   | 74  | 203   |    |    |        |     |     |     |
| Pitheciidae   | Callicebus | donacophilus | 909  |       |      |     |        |     |     |        |    |     |     |
## Appendix (continued)

<table>
<thead>
<tr>
<th>Family</th>
<th>Genus</th>
<th>Species</th>
<th>BM</th>
<th>AFR</th>
<th>GL</th>
<th>IBI</th>
<th>LS</th>
<th>NNM</th>
<th>WA</th>
<th>WM</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pitheciidae</td>
<td>Callicebus</td>
<td>hoffmannsi</td>
<td>1,030</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pitheciidae</td>
<td>Callicebus</td>
<td>moloch</td>
<td>956</td>
<td>3</td>
<td>164</td>
<td>12</td>
<td>1</td>
<td>74</td>
<td>60</td>
<td></td>
<td>12.0</td>
</tr>
<tr>
<td>Pitheciidae</td>
<td>Callicebus</td>
<td>personatus</td>
<td>1,380</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>15</td>
</tr>
<tr>
<td>Pitheciidae</td>
<td>Callicebus</td>
<td>torquatus</td>
<td>1,210</td>
<td>4</td>
<td>18</td>
<td>12</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>15</td>
</tr>
<tr>
<td>Pitheciidae</td>
<td>Chiropotes</td>
<td>albinasus</td>
<td>2,490</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>11.7</td>
</tr>
<tr>
<td>Pitheciidae</td>
<td>Chiropotes</td>
<td>satanas</td>
<td>2,580</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>15.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3,100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>24</td>
</tr>
<tr>
<td>Pitheciidae</td>
<td>Pithecia</td>
<td>aequatorialis</td>
<td>~2,250</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pitheciidae</td>
<td>Pithecia</td>
<td>irrorata</td>
<td>2,070</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pitheciidae</td>
<td>Pithecia</td>
<td>monachus</td>
<td>2,110</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>24.6</td>
</tr>
<tr>
<td>Pitheciidae</td>
<td>Pithecia</td>
<td>pithecia</td>
<td>1,580</td>
<td>2.0</td>
<td>164</td>
<td>19</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td>13.7</td>
</tr>
<tr>
<td>Atelidae</td>
<td>Alouatta</td>
<td>belzebul</td>
<td>5,520</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atelidae</td>
<td>Alouatta</td>
<td>caraya</td>
<td>4,330</td>
<td>3.7</td>
<td>187</td>
<td>1</td>
<td>262</td>
<td>325</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atelidae</td>
<td>Alouatta</td>
<td>palliata</td>
<td>4,020</td>
<td>3.5</td>
<td>186</td>
<td>20</td>
<td>1</td>
<td>320</td>
<td>325</td>
<td>1,100</td>
<td>13.0</td>
</tr>
<tr>
<td>Atelidae</td>
<td>Alouatta</td>
<td>pigra</td>
<td>6,430</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>~480</td>
</tr>
<tr>
<td>Atelidae</td>
<td>Alouatta</td>
<td>seniculus</td>
<td>4,670</td>
<td>4.5</td>
<td>191</td>
<td>17</td>
<td>1</td>
<td>295</td>
<td>372</td>
<td>630</td>
<td>25.0</td>
</tr>
<tr>
<td>Atelidae</td>
<td>Ateles</td>
<td>belzebuth</td>
<td>7,850</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atelidae</td>
<td>Ateles</td>
<td>chamek</td>
<td>9,330</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atelidae</td>
<td>Ateles</td>
<td>fusciceps</td>
<td>9,160</td>
<td>4.8</td>
<td>226</td>
<td>27</td>
<td>1</td>
<td>486</td>
<td></td>
<td></td>
<td>24.0</td>
</tr>
<tr>
<td>Atelidae</td>
<td>Ateles</td>
<td>geoffroyi</td>
<td>7,290</td>
<td>5.6</td>
<td>225</td>
<td>37</td>
<td>1</td>
<td>426</td>
<td>750</td>
<td>2,000</td>
<td>48.0</td>
</tr>
<tr>
<td>Atelidae</td>
<td>Ateles</td>
<td>paniscus</td>
<td>8,440</td>
<td>5</td>
<td>230</td>
<td>24</td>
<td>1</td>
<td>425</td>
<td>760</td>
<td>3,790</td>
<td>33.0</td>
</tr>
<tr>
<td>Atelidae</td>
<td>Brachyteles</td>
<td>arachnoides</td>
<td>8,070</td>
<td>7.5</td>
<td>233</td>
<td>34</td>
<td>1</td>
<td>638</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atelidae</td>
<td>Brachyteles</td>
<td>arachnoides</td>
<td>8,554</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Primate life histories**
<table>
<thead>
<tr>
<th>Family</th>
<th>Genus</th>
<th>Species</th>
<th>Population</th>
<th>Age</th>
<th>Sex</th>
<th>Body Mass</th>
<th>Life Span</th>
<th>Litter Size</th>
<th>Social Group Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atelidae</td>
<td>Lagothrix</td>
<td>lagothricha</td>
<td>5,585</td>
<td>5</td>
<td>223</td>
<td>24</td>
<td>1</td>
<td>432</td>
<td>315</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>7,020</td>
<td>7.58</td>
<td>225</td>
<td>1</td>
<td>447</td>
<td>507</td>
<td></td>
</tr>
<tr>
<td>Atelidae</td>
<td>Oreonax</td>
<td>flavicauda</td>
<td>~10,000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td>Allenopithecus</td>
<td>nigroviridis</td>
<td>3,180</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td>Miopithecus</td>
<td>talapoin</td>
<td>1,120</td>
<td>4.38</td>
<td>162</td>
<td>12</td>
<td>1</td>
<td>~178</td>
<td>180</td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td>Erythrocebus</td>
<td>patas</td>
<td>6,317</td>
<td>3</td>
<td>167</td>
<td>12</td>
<td>1</td>
<td>468</td>
<td>213</td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td>Chlorocebus</td>
<td>aethiops</td>
<td>2,980</td>
<td>4.75</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td>Cercopithecus</td>
<td>ascanius</td>
<td>2,920</td>
<td>5</td>
<td>172</td>
<td>12</td>
<td>1</td>
<td>336</td>
<td>201</td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td>Cercopithecus</td>
<td>campbelli</td>
<td>2,200</td>
<td>3.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td>Cercopithecus</td>
<td>cephus</td>
<td>2,805</td>
<td>5</td>
<td>170</td>
<td>11</td>
<td>1</td>
<td>339</td>
<td>366</td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td>Cercopithecus</td>
<td>diana</td>
<td>3,900</td>
<td>5.33</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td>Cercopithecus</td>
<td>erythrogaster</td>
<td>2,400</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td>Cercopithecus</td>
<td>erythrotis</td>
<td>~2,900</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>180</td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td>Cercopithecus</td>
<td>hamlyni</td>
<td>3,360</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td>Cercopithecus</td>
<td>lhoesti</td>
<td>3,450</td>
<td>5.42</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td>Cercopithecus</td>
<td>mitis</td>
<td>3,930</td>
<td>5.72</td>
<td>140</td>
<td>24</td>
<td>1</td>
<td>~337</td>
<td>692</td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td>Cercopithecus</td>
<td>mona</td>
<td>4,910</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td>Cercopithecus</td>
<td>neglectus</td>
<td>3,550</td>
<td>4.67</td>
<td>165</td>
<td>12</td>
<td>1</td>
<td>260</td>
<td>365</td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td>Cercopithecus</td>
<td>nictitans</td>
<td>4,216</td>
<td>4.9</td>
<td>170</td>
<td>11</td>
<td>1</td>
<td>406</td>
<td></td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td>Cercopithecus</td>
<td>petaurista</td>
<td>2,900</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td>Cercopithecus</td>
<td>pogonias</td>
<td>2,900</td>
<td>5</td>
<td>170</td>
<td>11</td>
<td>1</td>
<td>339</td>
<td></td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td>Cercopithecus</td>
<td>preussi</td>
<td>~4,500</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Appendix (continued)

<table>
<thead>
<tr>
<th>Family</th>
<th>Genus</th>
<th>Species</th>
<th>BM</th>
<th>AFR</th>
<th>GL</th>
<th>IBI</th>
<th>LS</th>
<th>NNM</th>
<th>WA</th>
<th>WM</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cercopithecidae</td>
<td><em>Cercopithecus</em></td>
<td>sclateri</td>
<td>~2,500</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td><em>Cercopithecus</em></td>
<td>solatus</td>
<td>3,920</td>
<td>4.75</td>
<td>18</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td><em>Cercopithecus</em></td>
<td>wolfi</td>
<td>2,870</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>~435</td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td><em>Macaca</em></td>
<td>arctoides</td>
<td>8,400</td>
<td>3.84</td>
<td>178</td>
<td>19</td>
<td>1</td>
<td>489</td>
<td>393</td>
<td>2,300</td>
<td>30.0²</td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td><em>Macaca</em></td>
<td>assamensis</td>
<td>6,900</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td><em>Macaca</em></td>
<td>cyclops</td>
<td>4,940</td>
<td>162</td>
<td>15</td>
<td>1</td>
<td>398</td>
<td>206</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td><em>Macaca</em></td>
<td>fascicularis</td>
<td>3,574</td>
<td>3.9</td>
<td>160</td>
<td>13</td>
<td>1</td>
<td>326</td>
<td>330</td>
<td>1,700</td>
<td>37.1¹⁸</td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td><em>Macaca</em></td>
<td>fuscata</td>
<td>8,030</td>
<td>5.54</td>
<td>173</td>
<td>24</td>
<td>1</td>
<td>503</td>
<td>365</td>
<td>2,730</td>
<td>33.0¹⁸</td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td><em>Macaca</em></td>
<td>maura</td>
<td>6,050</td>
<td>5</td>
<td>163²</td>
<td>22</td>
<td>1</td>
<td>389</td>
<td>500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td><em>Macaca</em></td>
<td>mulatta</td>
<td>5,370</td>
<td>3</td>
<td>165</td>
<td>12</td>
<td>1</td>
<td>466</td>
<td>192</td>
<td>1,454</td>
<td>36.0¹³</td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td><em>Macaca</em></td>
<td>nemestrina</td>
<td>4,900</td>
<td>3.92</td>
<td>167</td>
<td>14</td>
<td>1</td>
<td>444</td>
<td>234</td>
<td>1,320¹⁰</td>
<td>26.3²</td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td><em>Macaca</em></td>
<td>nigra</td>
<td>4,600</td>
<td>5.44</td>
<td>170</td>
<td>18</td>
<td>1</td>
<td>457</td>
<td></td>
<td></td>
<td>18.0²</td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td><em>Macaca</em></td>
<td>ochreata</td>
<td>2,600</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>~461</td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td><em>Macaca</em></td>
<td>radiata</td>
<td>3,700</td>
<td>4</td>
<td>162</td>
<td>12</td>
<td>1</td>
<td>388</td>
<td>365</td>
<td>2,000</td>
<td>30.0¹</td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td><em>Macaca</em></td>
<td>silenus</td>
<td>5,000</td>
<td>4.92</td>
<td>180</td>
<td>17</td>
<td>1</td>
<td>407</td>
<td>365</td>
<td></td>
<td>38.0¹</td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td><em>Macaca</em></td>
<td>sinica</td>
<td>3,200</td>
<td>5¹</td>
<td>18¹</td>
<td>1¹</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>30.0¹</td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td><em>Macaca</em></td>
<td>sylvanus</td>
<td>8,280</td>
<td>4.75</td>
<td>165</td>
<td>22</td>
<td>1</td>
<td>450</td>
<td>210</td>
<td>2,420</td>
<td>22.0¹</td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td><em>Macaca</em></td>
<td>thibetana</td>
<td>~11,000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Primate life histories
<table>
<thead>
<tr>
<th>Cercopithecidae</th>
<th>Species</th>
<th>Population</th>
<th>Mean Life Span</th>
<th>Median Life Span</th>
<th>Gestation</th>
<th>Weaning</th>
<th>Infant Mortality</th>
<th>Adult Mortality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Macaca tonkeana</td>
<td>12,800</td>
<td>~550</td>
<td>568</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mandrillus leucophaeus</td>
<td>8,450</td>
<td>722</td>
<td>28.6^2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mandrillus sphinx</td>
<td>11,350</td>
<td>613</td>
<td>348</td>
<td>3,000</td>
<td>29.1^2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cercocebus albigena</td>
<td>6,209</td>
<td>500^3</td>
<td>365</td>
<td>2,170</td>
<td>32.7^1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cercocebus agilis</td>
<td>5,660</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cercocebus atys</td>
<td>6,200</td>
<td>316</td>
<td>26.8^13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cercocebus gasteritus</td>
<td>5,260</td>
<td>171</td>
<td>19.0^2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cercocebus torquatus</td>
<td>5,500</td>
<td>171</td>
<td>27.0^1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lophocebus albigena</td>
<td>6,020</td>
<td>186</td>
<td>220</td>
<td>2,200^20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lophocebus aterrimus</td>
<td>5,760</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Papio anubis</td>
<td>11,700</td>
<td>915</td>
<td>3,800^3</td>
<td>25.2^13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Papio cynocephalus</td>
<td>9,750</td>
<td>710</td>
<td>2,500</td>
<td>40.0^1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Papio hamadryas</td>
<td>9,900</td>
<td>695</td>
<td>3,100</td>
<td>37.5^13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Papio papio</td>
<td>12,100</td>
<td>604</td>
<td>40.0^1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Papio ursinus</td>
<td>14,773</td>
<td>600</td>
<td>45.0^1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Theropithecus gelada</td>
<td>11,427</td>
<td>465</td>
<td>3,900</td>
<td>28.0^13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Procolobus verus</td>
<td>4,200</td>
<td>170</td>
<td>24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Procolobus badius</td>
<td>7,421</td>
<td>25</td>
<td>1</td>
<td>522</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colobus angolensis</td>
<td>7,570</td>
<td></td>
<td></td>
<td>22.3^19</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Appendix (continued)

<table>
<thead>
<tr>
<th>Family</th>
<th>Genus</th>
<th>Species</th>
<th>BM</th>
<th>AFR</th>
<th>GL</th>
<th>IBI</th>
<th>LS</th>
<th>NNM</th>
<th>WA</th>
<th>WM</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cercopithecidae</td>
<td>Colobus</td>
<td>guereza</td>
<td>7,900</td>
<td>4.75</td>
<td>170</td>
<td>20</td>
<td>1</td>
<td>445</td>
<td>330</td>
<td>1,600</td>
<td>24.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>8,102</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>549</td>
<td>390</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9,200</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>394</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td>Colobus</td>
<td>polykomos</td>
<td>7,662</td>
<td>8.5</td>
<td>170</td>
<td>22</td>
<td>1</td>
<td>400</td>
<td>215</td>
<td>1,240</td>
<td>30.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>8,300</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>480</td>
<td>394</td>
<td></td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td>Colobus</td>
<td>satanas</td>
<td>7,420</td>
<td></td>
<td></td>
<td>195</td>
<td>1</td>
<td></td>
<td>500</td>
<td>219</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>480</td>
<td>394</td>
<td></td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td>Colobus</td>
<td>vellerosus</td>
<td>6,900</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td>Semnopithecus</td>
<td>entellus</td>
<td>6,910</td>
<td>3.42</td>
<td>184</td>
<td>17</td>
<td>1</td>
<td>500</td>
<td>249</td>
<td>2,100</td>
<td>25.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9,890</td>
<td>3.9</td>
<td>200</td>
<td></td>
<td></td>
<td></td>
<td>396</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>10,280</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>416</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>14,800</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td>Semnopithecus</td>
<td>vetulus</td>
<td>5,797</td>
<td>4</td>
<td>200</td>
<td>32</td>
<td>1</td>
<td>360</td>
<td>219</td>
<td>1,100</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5,900</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>365</td>
<td></td>
<td>31.1</td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td>Trachypithecus</td>
<td>cristata</td>
<td>5,760</td>
<td>4</td>
<td></td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>6,300</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>31.1</td>
<td></td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td>Trachypithecus</td>
<td>francoisi</td>
<td>7,300</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>457</td>
<td>394</td>
<td></td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td>Trachypithecus</td>
<td>geei</td>
<td>9,500</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td>Trachypithecus</td>
<td>johnii</td>
<td>11,200</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td>Trachypithecus</td>
<td>obscura</td>
<td>6,260</td>
<td>4</td>
<td>150</td>
<td>2</td>
<td>1</td>
<td>341</td>
<td>365</td>
<td></td>
<td>25.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>6,530</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>485</td>
<td>25.0</td>
<td></td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td>Trachypithecus</td>
<td>phayrei</td>
<td>6,300</td>
<td>165</td>
<td>15</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>305</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>10,500</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td>Trachypithecus</td>
<td>pileatus</td>
<td>9,860</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>458</td>
<td></td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td>Presbytis</td>
<td>comata</td>
<td>6,710</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td>Presbytis</td>
<td>frontata</td>
<td>5,670</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td>Presbytis</td>
<td>hosei</td>
<td>5,630</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td>Presbytis</td>
<td>melalophos</td>
<td>6,470</td>
<td>3.8</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>16.0</td>
<td></td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td>Presbytis</td>
<td>potenziani</td>
<td>6,400</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td>Presbytis</td>
<td>rubicunda</td>
<td>6,170</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td>Presbytis</td>
<td>senex</td>
<td>5,797</td>
<td>3.8</td>
<td>195</td>
<td>20</td>
<td>1</td>
<td>360</td>
<td>214</td>
<td>1,100</td>
<td>23.0</td>
</tr>
</tbody>
</table>

**Note:** BM = body mass (kg), AFR = age at first reproduction, GL = gestation length (days), IBI = interbirth interval (days), LS = lifespan (years), NNM = number of newborns, WA = weight of adult female (kg), WM = weight of male (kg), L = life span (years).
<table>
<thead>
<tr>
<th>Family</th>
<th>Species</th>
<th>Scientific Name</th>
<th>Males</th>
<th>Females</th>
<th>Litter Size</th>
<th># Pregnancies</th>
<th># Infants</th>
<th>Frequency</th>
<th>Longevity</th>
<th>Males</th>
<th>Females</th>
<th>Lifespan</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cercopithecidae</td>
<td><em>Presbytis</em> thomasi</td>
<td>6,690</td>
<td>6,690</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.13</td>
<td>281</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td><em>Nasalis</em> larvatus</td>
<td>9,593</td>
<td></td>
<td></td>
<td>1</td>
<td>450</td>
<td>210</td>
<td>200</td>
<td>21.0^13</td>
<td>281</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td><em>Simias</em> concolor</td>
<td>6,800</td>
<td></td>
<td></td>
<td>1</td>
<td>463</td>
<td></td>
<td>10.3^1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td><em>Pygathrix</em> nemaeus</td>
<td>8,180</td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>463</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td><em>Rhinopithecus</em> avunculus</td>
<td>8,000</td>
<td></td>
<td></td>
<td>1</td>
<td>463</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td><em>Rhinopithecus</em> bieti</td>
<td>9,960</td>
<td></td>
<td></td>
<td>1</td>
<td>463</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cercopithecidae</td>
<td><em>Rhinopithecus</em> roxellana</td>
<td>11,600</td>
<td></td>
<td></td>
<td>1</td>
<td>463</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hylobatidae</td>
<td><em>Hylobates</em> agilis</td>
<td>5,820</td>
<td>5,820</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>12</td>
<td>1</td>
<td>724^12</td>
<td>700^2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hylobatidae</td>
<td><em>Hylobates</em> color</td>
<td>5,749</td>
<td>5,749</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>12</td>
<td>1</td>
<td>724^12</td>
<td>700^2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hylobatidae</td>
<td><em>Hylobates</em> hoolock</td>
<td>6,880</td>
<td>6,880</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>12</td>
<td>1</td>
<td>724^12</td>
<td>700^2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hylobatidae</td>
<td><em>Hylobates</em> klossii</td>
<td>5,920</td>
<td>5,920</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>12</td>
<td>1</td>
<td>724^12</td>
<td>700^2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hylobatidae</td>
<td><em>Hylobates</em> lar</td>
<td>5,340</td>
<td>5,340</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>12</td>
<td>1</td>
<td>724^12</td>
<td>700^2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hylobatidae</td>
<td><em>Hylobates</em> leucogenys</td>
<td>7,320</td>
<td>7,320</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>12</td>
<td>1</td>
<td>724^12</td>
<td>700^2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hylobatidae</td>
<td><em>Hylobates</em> moloch</td>
<td>5,292</td>
<td>5,292</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>12</td>
<td>1</td>
<td>724^12</td>
<td>700^2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hylobatidae</td>
<td><em>Hylobates</em> muelleri</td>
<td>5,350</td>
<td>5,350</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>12</td>
<td>1</td>
<td>724^12</td>
<td>700^2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hylobatidae</td>
<td><em>Hylobates</em> pileatus</td>
<td>5,440</td>
<td>5,440</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>12</td>
<td>1</td>
<td>724^12</td>
<td>700^2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hylobatidae</td>
<td><em>Hylobates</em> syndactylus</td>
<td>10,568</td>
<td>10,568</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>12</td>
<td>1</td>
<td>724^12</td>
<td>700^2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pongidae</td>
<td><em>Pongo</em> pygmaeus</td>
<td>35,700</td>
<td>35,700</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>12</td>
<td>1</td>
<td>724^12</td>
<td>700^2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pongidae</td>
<td><em>Pongo</em> abelii</td>
<td>35,600</td>
<td>35,600</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>12</td>
<td>1</td>
<td>724^12</td>
<td>700^2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pongidae</td>
<td><em>Gorilla</em> gorilla</td>
<td>71,000</td>
<td>71,000</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>12</td>
<td>1</td>
<td>724^12</td>
<td>700^2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pongidae</td>
<td><em>Pan</em> paniscus</td>
<td>33,200</td>
<td>33,200</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>12</td>
<td>1</td>
<td>724^12</td>
<td>700^2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pongidae</td>
<td><em>Pan</em> troglodytes</td>
<td>30,000</td>
<td>30,000</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>12</td>
<td>1</td>
<td>724^12</td>
<td>700^2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix (continued)

<table>
<thead>
<tr>
<th>Family</th>
<th>Genus</th>
<th>Species</th>
<th>BM</th>
<th>AFR</th>
<th>GL</th>
<th>IBI</th>
<th>LS</th>
<th>NNM</th>
<th>WA</th>
<th>WM</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hominidae</td>
<td>Homo</td>
<td>sapiens</td>
<td>45,800</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>42,200</td>
<td>14</td>
<td>267</td>
<td>36</td>
<td>1</td>
<td>2,900</td>
<td>730</td>
<td>10,980</td>
<td>120.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>45,800</td>
<td>15</td>
<td>270</td>
<td></td>
<td></td>
<td>3,334</td>
<td>930</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>73,200</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For each currently recognized primate species, principal life history data are presented. One should note the large number of empty cells and the variation in some variables used in the literature. All values without superscript are from Kappeler and Pereira (2003). Other sources are listed below.


BM (Body mass) = mean body mass (g) of wild adult females, ~ indicates mixed or unknown sex sample; AFR (Age at first reproduction) = mean age (years) of first female reproduction, GL (Gestation length) = mean gestation length (days), IBI (Interbirth interval) = mean interbirth interval (month), LS (Litter size) = modal litter size, NNM (Neonatal mass) = mean body mass (g) of neonate females, ~ indicates mixed or unknown sex sample, WA (Weaning age) = mean age at weaning (days), WM (Weaning mass) = mean body mass (g) at weaning, L (Longevity) = maximal recorded life span in years.
References


Charnov EL, Berrigan D (1993) Why do female primates have such long lifespans and so few babies? or life in the slow lane. Evol Anthropol 1: 191–194


Cheney DL, Seyfarth RM, Andelman SJ, Lee PC (1988) Reproductive success in vervet...


Joffe TH (1997) Social pressures have selected for an extended juvenile period in primates. J Hum Evol 32: 593–605


11 The Biology and Evolution of Ape and Monkey Feeding

Joanna E. Lambert

Abstract
Like all animals, primates must garner sufficient energy and nutrients from their habitat to accomplish other biological imperatives such as mating and avoiding predators. The order Primates exhibits an extraordinary diversity of feeding and foraging-related adaptations to meet this imperative, some of which are shared with other taxa, others of which are unique to primates. In this chapter, I explore the evolutionary underpinnings of these adaptations as well as evaluate their ecological implications. I first discuss several unique aspects of primate feeding biology, including the evolution of large brains, trichromatic color vision, and tool use. I then move on to evaluate the fundamental problems of plant fiber and chemical defenses and conclude by arguing that primate adaptations for fermenting fiber and detoxifying plant chemical defenses have implications for primate species adaptations, density, and diversity since the Miocene. Because of the implications for understanding adaptations in our own lineage, in this latter section I pay particular attention to African primates; I employ a strongly comparative approach and evaluate hominoid feeding biology in light of what we know about Cercopithecoida.

11.1 Introduction

Feeding is a biological imperative. Like all animals, primates must garner sufficient energy and nutrients from their habitat to accomplish other biological imperatives such as mating and avoiding predators. This rather obvious fact has not gone unnoticed by primate biologists and inquires into how and what primates consume have been central to primate studies since the 1970s (Clutton-Brock 1977; Chivers et al. 1984; Rodman and Cant 1984; Altmann 1998). Most primates are omnivorous and, by definition, consume both plant and animal matter. Nonetheless, primate requirements for energy and nutrients are met...
largely by plants (Milton 1993). The system of energy transfer from plants to animals via herbivory evolved during the Permian, ca. 255 Ma. This is the food chain with which we are most familiar [i.e., producer (plants) → primary consumer (herbivores) → secondary consumer (carnivores)]; the addition of an herbivore trophic level represents a marked shift from previous time periods that lacked primary consumers (Potts 1996). This changed dramatically throughout the Mesozoic as Angiosperms (flowering plants) evolved, and as amphibians, reptiles, and eventually mammals radiated into a diversity of plant-consuming niches. Primates is thus among these modern Orders that evolved to feed primarily from the Kingdom Plantae, and they do so with staggering diversity in their foraging and feeding adaptations. Indeed, extant primates forage in literally every possible social permutation: as solitary animals, as mated pairs, in feeding parties that fission and fuse, and in cohesive social groups (Strier 2003). Primates procure food in a diversity of ways as well, and there is virtually nothing that some primate will not eat, from fungi and lichen, grass and palm nuts, nectar and termites, snakes and antelope; accordingly, primates exhibit a diversity of morphological adaptations in dentition, digestion, locomotor apparatus, and myriad cranial and postcranial features.

Evolution is a wondrous thing, resulting in adaptations that are unique to species (autapomorphies), shared by several related taxa (usually synapomorphies), or inherited from an original, ancestral state (plesiomorphy). Moreover, “natural selection, acting in various guises at various levels, seems together with genetic drift to account for almost all features of organisms once the appropriate raw material has arisen by mutation and recombination” (Futuyma 1979 p 438). Hence, we can assume that feeding-related features observed in primate species today are (1) the result of natural selection and other microevolutionary forces and are either (2) unique to a species (e.g., mandible morphology), shared by a subset of primate taxa (e.g., cercopithecine cheek pouches), shared by all primates (e.g., postorbital bar), or are ancestral to mammals (e.g., heterodonty), vertebrates (e.g., protected spinal cord), or animals generally (e.g., reliance on ingestion). Given this extraordinary diversity at multiple scales, in this chapter I cannot do justice to the entire array of strategies employed by primate species for procuring nutrients from their environments (for more information, see Garber 1987; Milton 1993; Janson and Chapman 1999; Lambert in press). Instead, my focus here is to first evaluate several noteworthy—indeed, unique to the Order—aspects of primate feeding biology. Thus, in the first part of this chapter, I evaluate the feeding implications of large brains, trichromatic color vision, and tool use in primates. I next discuss the fundamental problem, encountered by all plant-consuming animals in general, of plant fiber and chemical defenses. I conclude by arguing that primate adaptations for fermenting fiber and
detoxifying plant chemical defenses have had implications for primate species adaptations, density, and diversity since the Miocene. While primates undoubtedly encounter such feeding challenges in all biogeographic areas in which they are found, because of the implications for understanding adaptations in our own lineage I pay particular attention to Africa. I employ a strongly comparative approach and evaluate hominoid feeding biology in light of what we know about Cercopithecoidea.

11.2 Feeding adaptations unique to primates: evolution of large brains, color vision, and tool use

Roughly three-fourths of all primate species forage and feed in cohesive social groups. The observed size of a primate social group is often argued to be one that is both big enough to provide protection from predators and small enough that group members do not incur too great a cost from increasing feeding competition (Struhsaker 1981; Terborgh and Janson 1986; Dunbar 1988; Garber 1988; Janson and van Schaik 1988; van Schaik 1989; Chapman 1990; Isbell 1991; Janson 1992; van Schaik and Kappeler 1993; Chapman et al. 1995; Janson and Goldsmith 1995; Treves and Chapman 1996; Sterck et al. 1997; Boinski et al. 2000; Gillespie and Chapman 2001; Chapman and Chapman 2002; Koenig 2002). As a consequence of extreme variability in food availability in different habitats, primate species and populations experience the cost of feeding competition to varying degrees. Yet, regardless of the degree to which primates do or do not compete for food, they must meet their basic nutritional needs for macronutrients (carbohydrates, protein, lipids) and micronutrients (vitamins, minerals, and water). Primates require the full suite of nutrients needed by mammals in general (45–47 in total of amino acids, fatty acids, vitamins, and minerals); large quantities of macronutrients are required for energy related to growth and maintenance, while micronutrients are vitally important for innumerable physiological processes (Oftedal and Allen 1996; Leonard 2000). In meeting these needs, primates are confronted with an array of feeding challenges that influence food quality. These challenges can be intrinsic, representing some inherent chemical, nutritional, or structural feature of the food—or extrinsic—a function of availability of that food and the costs (e.g., increased ranging, competition) associated with feeding on it as a consequence of that availability (Lambert in press). Primates are not unique in having to deal with feeding challenges, but several characteristically primate feeding adaptations are particularly deserving of attention for their extreme utility in feeding: brain expansion, trichromatic color vision, and tool use.
11.2.1 Brains

“The most distinctive characteristic of primates is the size and complexity of the brain” (Napier and Napier 1985 p 34). In making this comment, the Napiers echoed the work of Le Gros Clark (1959) who argued that the extreme neocortical expansion observed in modern humans is a continuation of a trend in mammals in general (whose brains are on average $10 \times$ larger than those of reptiles), and primates in particular. Indeed, primates as an Order have a brain that is $2 \times$ larger than what would be expected of mammals of a particular body size, apes roughly $3–4 \times$, and modern humans $7–8 \times$ (Jerison 1973; Martin 1996; Lewin and Foley 2004). The increase in brain size is particularly notable in the neocortex and especially in those areas of the brain concerned with vision and higher functioning (Radinsky 1977; Martin 1990). As such, discussions regarding primate evolution generally—and human evolution specifically—almost invariably refer to brain size expansion, if they do not completely revolve around it. Primate biologists have thus devoted much time identifying biological and ecological variables that may have been of particular selective importance in the evolution of brain size. Given the fact that primates are also among the most social of mammals, a number of hypotheses regarding brain expansion have centered on the cognitive demands of negotiating and maintaining complicated social relationships (Byrne and Whiten 1988; Barton 2000). Others, however, are focused on the relationships among brain size, metabolism, and dietary quality, while still others have evaluated the role of foraging in ecologically complex landscapes in selecting cognitive capacity (Milton 1988; Aiello and Wheeler 1995; Garber 2000).

Brain tissue is metabolically expensive, and there are differences in relative brain size among primate species (Aiello and Wheeler 1995). Interestingly, there is no correlation between relative brain size and basal metabolic rates in mammals. It has thus been proposed that the metabolic expense of maintaining a large brain in some primate species is offset by a reduction in other expensive tissues, especially the digestive tract (Aiello and Wheeler 1995). Total volume and mass of the gastrointestinal tract is reasonably predictable of the food type—and its quality—that primates consume. Chivers and Hladik (1980), for example, determined that mammals relying primarily on vertebrate prey have a simple stomach, an unspecialized colon, and a long, complex small intestine. In contrast, species relying heavily on vegetative plant parts are typically characterized in having large chambers (stomach, caecum, colon) in the gut for fermenting structural carbohydrates (fiber), while mammals that rely on fruit are noted for their relatively unspecialized gut. Differences in gut morphology among frugivores are generally attributed to the degree to which each species relies on either insects or leaves for protein to supplement its fruit diet.
Relatively more encephalized primate species tend also to be those primates species with relatively smaller and simpler guts and those that consume a higher-quality diet (Milton 1988; Aiello and Wheeler 1995). Quality in this context is best defined in terms of the food’s capacity to yield energy. Foods differ widely in this respect, and some require more digestive processing (either by endogenous enzymes or by microbial fermentation) than others before energy is yielded (Lambert 1998). Fruit, for example, while exhibiting wide interspecific variation, is generally a source of easily digested, nonstructural carbohydrates (esp. fructose), while leaves tend to be high in structural polysaccharides (fiber) and require more time and digestion before energy is yielded. Insects (both mature and larval) can be a good source of protein and lipids, although the costs of searching and handling time and the polysaccharides of chitonous exoskeletons mitigate their overall quality. Meat from vertebrate animals is an excellent source of protein and fats and is easily digested (by the action of pepsin, trypsin, and chymotrypsin in the stomach; Tortora and Anagnostakos 1987). Thus, the concomitant facts that modern humans have the largest relative brain size among animals, a simple and relatively small total gut for their body size, and an evolutionary history of meat-consumption, is of considerable interest to paleoanthropologists and primate feeding biologists alike. Indeed, with regard to human evolution, Milton (1999) has argued that: “by routinely including animal protein in their diet, they [i.e., humans, my insert] were able to reap some nutritional advantages enjoyed by carnivores, even though they have feature of gut anatomy and digestive kinetics of herbivores” (p 11). In other words, by increasingly including meat (a readily digested, high-quality food) in the diet, our human ancestors garnered the requisite energy for maintaining an increase in body size, while at the same time evolving a larger brain at the expense of total gut size (Aiello and Wheeler 1995; Milton 1999).

There thus appears to be a relationship between relative brain size, relative gut size, and the quality of food. Some scholars have also noted the relationships among brain size, cognitive capacity, and food distribution and availability. As mentioned, primates consume a broad diversity of food types, largely of plant origin, and to a lesser degree of invertebrate and vertebrate animals. For ease of discussion, plant foods are often lumped into broad feeding categories, including vegetative plant parts (e.g., leaves) and reproductive plant parts (e.g., fruit, seeds, flowers). Vegetative and reproductive plant foods in a habitat are distributed neither evenly nor equally in space and time, and there is wide variation within and between habitats and species in their richness, abundance, and nutritional quality. Thus, the initial challenge to all animal consumers is finding food in the first place: not an insignificant task (Oates 1987; White 1998). “Typical” habitats do not exist, and generally speaking, primate foods resources are difficult to find
(Milton 1988; Janson and Chapman 1999). Indeed, in an analysis of tree abundance in Barro Colorado Island, Panama, Milton (1988) found that 65% of all tree species were encountered less than once per hectare, and ripe fruits were available for individual trees only 0.8 months/annum. Nonetheless, there are gross differences in availability among broad food types and, for example, within a given habitat vegetative plant parts such as mature leaves tend to be more abundant and predictably available relative to reproductive plant parts such as fruit, seeds, flowers, and nectar which are more patchy in space and time (Milton 1980; Isbell 1991). This is a function of the fact that most tropical forests are characterized by tree species with differences in seasonal and annual fruit production and by trees that either produce fruit in small quantities or produce abundant fruit in widely scattered trees (Janzen 1967; Frankie et al. 1974; Whitmore 1990; van Schaik et al. 1993).

The primate foods that differ in their distribution and availability suggest different selective pressures with regard to cognitive capacity: some foods are essentially more difficult to monitor and locate than others (Garber 2004). Parker and Gibson (1977), for example, suggested that large brain size correlates with omnivorous feeding in primates. Milton’s (1988) long-term research on *Alouatta palliata* (mantled howler monkeys) and *Ateles geoffroyi* (spider monkey) identified important differences in foraging and feeding that also correspond to brain size. The relatively small-brained howler monkey, while at times of the year highly frugivorous, in general relies heavily on leaves, which are both more abundantly distributed and more refractory to digestion than fruit. During some seasons, howler monkeys can spend upward of 90% of daily feeding time on leaves. Conversely, the relatively larger-brained spider monkey is highly frugivorous and spends an annual average of 72% daily feeding time on fruit. The overall differences in time devoted to fruit versus leaves and in distribution and availability of vegetative and reproductive plant parts are reflected in total home range size of the two species: 31 hectares in howler monkeys, and 800 hectares in spider monkeys. Thus, argues Milton (1988 p 289): “spider monkeys are faced with a far more complex problem than howlers with respect to locating their food sources since, in effect, they are dealing with a supplying area over 25 times as large.” A more patchily distributed food source requires greater cognitive capacities for dealing with a more complex ecological matrix. Spider monkeys exhibit a relative brain size and degree of neural complexity approximately $2 \times$ that of howler monkeys (*Ateles* body size: 7.6; brain weight: 107 g; *Alouatta* body size: 6.2 kg; brain weight: 28 g). And while it is recognized that indices of relative brain size are not necessarily direct indicators of cognitive capacity, in this case Milton (1988) argues that it does appear that selective pressure for spatial memory is more
intense for spider monkeys (a frugivore) than howler monkeys (a folivore), which too may help to explain differences in gross brain size and complexity.

Building on these ideas, most recently Potts (2004) has proposed a “fruit-habitat hypothesis” that suggests the evolution of the relatively very large brains of great apes, and their concomitant cognitive capacity, can be explained by a “…causal connection between ape ancestral diets, habitats, and environmental history…” (p 224). In short, he argues that as preferred fruit resources became increasingly rare as a function of forest reductions and climatic shifts in the African Miocene, ancestral apes were under extreme selective pressure for evolving cognitive means (e.g., complex mental representational ability) to deal with food source uncertainty.

Other feeding related features have been linked to relative brain size, such as extractive foraging (Barton 2000). For example, members of the most encephalized platyrrhine genus, Cebus, are highly manipulative, destructive foragers, and are highly adept at extracting resources from tree cavities, as well as difficult to penetrate food types such as hard palm nuts. The aye-aye (Daubentonia madagascariensis) is among the most encephalized prosimians and a majority of its calories are derived from arthropod grubs that they have extracted deep from within branches.

11.2.2 Vision

Primates exhibit a diversity of visual adaptations that distinguish them from other mammals, several of which have been related directly to feeding and foraging. Primates, for example, have very large eyes relative to body size and a high degree of orbital convergence; both traits are argued to be of selective advantage for nocturnal visual predation (Cartmill 1992; Ross 2000). Primates, too, are unique among mammals in having trichromatic color vision. Color vision relies on the presence of photoreceptors that contain light-sensitive opsin proteins (cones). Most vertebrates (including reptiles, birds, and some fish) have four such photoreceptors (hence, “tetrachromacy”), each with different peak light spectral sensitivity (Ross 2000; Dominy 2004; Dominy et al. 2004). However, early in the radiation of mammals at least two of these photoreceptors were lost, the legacy of which is that most extant Eutherian mammals have only two opsin proteins, resulting in dichromatic color vision. This is often explained by the fact that the earliest niche occupied by mammals was a nocturnal one; since cones perform poorly under low-light conditions, the selection for their maintenance was relaxed. It is thus of considerable interest that in Primates trichromacy has not
only evolved once, but potentially several times. The primitive mammalian condition includes one short-wave (430 nm, “S”) and one long-wave (560 nm “L”) sensitive cone pigment, with the long-wave pigment gene located on the X-chromosome. Subsequent gene duplications on the X-chromosome in catarhines resulted in separate middle-wave (530 nm, “M”) and long-wave pigments; all cercopithecoids and hominoids tested to date thus exhibit the three cone opsins. Platyrrhines are highly variable in their perception of color, since there is a single X-chromosome locus for the opsin gene. With the exception of Aotus (a monochromat) and Alouatta (routinely trichromatic), all tested platyrrhines ceboids are polymorphic at the locus for the long-wave pigment and thus have mixed populations of dichromats (males and homozygous females) and trichromats (heterozygous females) (Ross 2000; Dominy 2004). Although the gene duplication in Alouatta is similar to that of Catarrhini, it is an independent evolutionary event, and probably occurred quite recently (Jacobs 1996; Dominy 2004).

The combination of three opsins in the retina being tuned to different wavelengths facilitates perception of an enormous range of color hues. The fact that foods in natural habitats exhibit an equally large range of color has resulted in several foraging-related hypotheses for the explanation of the evolution of color vision in primates. Generally, these hypotheses emphasize the advantages of color vision in detecting fruit, leaves, or both (Ridley 1894; Snodderley 1979; Lucas et al. 1998; Dominy 2004). For example, since at the least the turn of the century, it has been suggested that color vision is linked to the evolution of colorful fruit in angiosperms (Ridley 1894; Polyak 1957; Snodderley 1979; Regan et al. 1998). In this putatively coevolved relationship, flowering tree species evolved colorful fruit to ensure the attraction of seed dispersers (such as primates) and (some) primates became better at finding such fruit as a result of their trichromatic color vision. There is no doubt that primates are important seed dispersers; as a taxon they are often the most abundant arboreal frugivores in tropical forests (Lambert 2002a; Lambert and Chapman 2005). Whether primate feeding has selected for fruit color and whether color vision coevolved with frugivory is much less clear. “Coevolution” is a narrowly defined event, wherein there is: “… an evolutionary change in a trait of the individuals in one population in response to a trait of the individuals of a second population, followed by an evolutionary response to the change in the first” (Janzen 1980 p 611). As such, coevolution represents a highly specialized and interdependent step-by-step process of interrelated evolutionary change in two or more species and differs from both adaptation and exaptation in terms of the duration, strength, and mutual historical relationships between the interacting species. In a detailed analysis
of both catarrhine and platyrrhine frugivory and seed-handling, Lambert and Garber (1998) found that primates in both allopatry and sympatry exploit a wide variety of plant species exhibiting a very broad and unpredictable array of fruit colors. Moreover, across taxa, these researchers found no evidence for any particular or consistent suite of fruit traits or patterns in the way fruits are processed and dispersed. Consistent with these findings, Dominy (2004) determined that primate frugivory is highly generalized with respect to hue and argues that rather than fruit color, it is either abundance or accessibility (or both) that are the more important determinants of fruit choice by primates. Moreover, the relationship between fruit color and nutritional properties (especially in terms of the sugar “reward”) is equivocal; hue and sugar are uncorrelated in Costa Rica and Peru (Wheelwright and Janson 1985; Dominy 2004). Nonetheless, the “... old idea that primate trichromacy evolved in the context of fruit detection and identification enjoys some current support” (Jacobs 1996 p 198).

More recently, Lucas et al. (1998) proposed an alternative hypothesis imputing leaf-selection, rather than fruit, as the important agent of evolution for trichromacy. As demonstrated by these researchers, leaf toughness is highly correlated with leaf color; toughness is also positively correlated with high fiber and negatively correlated with protein. These authors thus suggest that trichromatic vision may provide an important advantage in finding palatable and nutritious leaves in the tropics where many tree species delay the “greening” of leaves in an apparent attempt to ward off herbivores. In a study that focused on two colobine species, one cercopithecine, and one ape (Pan troglodytes) in the Kibale National Park, Uganda, Dominy and Lucas (2001) demonstrated that increasing reddishness of edible leaves is significantly correlated with reduced toughness and protein. Such a relationship was not found with fruit, and the fact that Alouatta—a folivore—is the only routine trichromat in the New World generally lends support to leaf-related hypothesis. The evolutionary context for trichromacy and its function in leaf-detection may stem to early Anthropoidea. Climatologically, the Eocene–Oligocene transition is generally marked by dramatic cooling, which, in addition to many other ecological shifts, resulted in extinction or reduction in palms in Africa and Asia, although not in Madagascar or South America (Morley 2000; Dominy 2004). Dominy (2004) suggests with the loss of this critical keystone resource (i.e., palms), stem catarrhines turned to protein-rich young leaves; a chance mutation on the duplication of an opsin gene would have provided the underlying visual mechanism to exploit this new resource effectively. Primates in South America or Madagascar would not have been under such intense selection, given the availability of palms as an important resource.
11.2.3 Tool use

Several animal species (e.g., sea otters, woodpecker finches) have been noted for their use of objects in their environment and hence meet the definition of tool use: “... the employment of an unattached environmental object to alter more efficiently the form, position, or condition of another object, organism, or the user itself” (Beck 1980 p 10). Yet tool use is rare in nature (Garber 2004), and tool use in nonhuman primates differs from that of other animals in several respects, including the long period of learning, the complexity of skills, the imputed cognitive abilities, and in the case of common chimpanzees (P. troglodytes), the diversity of tool kits and use of tool-composites (Strier 2003). Nonhuman primates do not, however, make tools to make tools; this accomplishment appears to be uniquely human (McGrew 1992).

Besides Homo sapiens and common chimpanzees, tool use among primates has been observed in capuchin monkeys (Cebus), organgutans (Pongo), western gorillas (Gorilla), and bonobos (Pan paniscus). In most cases, primate tool use is undertaken within the context of feeding, the most famous example coming from Gombe National Park, Tanzania, where in 1960 Jane Goodall first observed and described chimpanzees extracting termites (Macrotermes bellicosus) from their mounds by using modified plant stems:

Passages into a termite mound are narrow and not completely straight, so the materials used must be smooth and fairly pliable if they are to be effective. Tools are fashioned from grasses, vines, bark, twigs, or palm frondlets. Sometimes a chimpanzee will pick up almost any suitable material that is nearby, including the discard tools of others who have worked the mound previously. At other times clumps of grass, tangles of vines, and so on are carefully inspected before a tool is selected; a length may be picked, then discarded immediately before it has been used, and another choice made. To some extent the procedure reflects individual differences, but dry-season termiting call for more skill and more care in the choice of material that wet-season fishing, when (a) the insects are near the surface and (b) the soldiers are on the defensive and quick to bite at any foreign material inserted into the nest.” (Goodall 1986 p 536)

Pongo pygmaeus, too, has been observed to extract social insects (e.g., ants, termites) from crevices and holes in trees, as well as use twigs to pry seeds from Neesia spp. fruit (van Schaik and Fox 1996). In addition to fishing for termites and other social insects, the best known feeding-related tool use in chimpanzees includes hammer and anvil technology for cracking open nuts. This form of tool use, wherein a stone or log is used as an anvil in combination with a rock hammer, is generally viewed as the most complex form of tool use in any nonhuman
animal (Fragaszy et al. 2004). It is both a tool-composite (“... two or more tools having different functions that are used sequentially and in association to achieve a single goal”; Sugiyama 1997 p 25) and it involves employing multiple spatial relations in sequence.

Marked contrasts in tool use and tool kits are found across populations of chimpanzees in Africa; some of these differences are clearly cultural, resulting from social traditions rather than ecological differences or availability of materials (McGrew 1992; Whiten et al. 1999; Whiten et al. 2001). Hammer and anvil technology, for example, appears to be uniquely West African. Although tool use by nonhuman primates, especially chimpanzees, differs in terms of what the tool is made of, how they are made, the contexts they are used in, and the tasks they are used for, a picture of their utility in dealing with food scarcity is beginning to emerge. As discussed previously, fruit as a primate food resource is particularly patchy both spatially and temporally. As it happens, the two primate taxa for which tool use is commonly reported (i.e., *P. troglodytes* and *Cebus* spp.) are also among the most highly frugivorous of primate taxa. Fruit is an ephemeral resource and frugivorous animals must have strategies for dealing with those times of the year when their preferred foods are not available. In many cases, primate species fall back on alternate foods (e.g., bark, leaves, terrestrial herbaceous vegetation). If these fallback resources are critical to sustaining population size (i.e., truly “limiting”) they then meet the definition for “keystone” resources which are typically consumed in quantities that compensate for the scarcity of the animals’ main foods (Gilbert 1980; Terborgh 1986). Several authors have recently suggested that the use of tools during such periods may facilitate the use of fallback, indeed keystone, foods not otherwise available to them (Yamkoshi 1998; Fragaszy et al. 2004; Moura and Lee 2004).

The habituated chimpanzee community at Bossou, Guinea, lives in both a very small (5 km²) and isolated forest, leaving this community without the ability to either shift food types or expand their range in search of fallback foods (Yamkoshi 1998). Overall resource richness is low, and seasonality of fruit availability in this region of West Africa is extreme. Chimpanzees in this area employ a number of feeding-related tools, including hammer and anvil, pestle-pounding, ant-dipping, and algae-scooping with leaves. Tool use in these chimpanzees facilitates the consumption of two important keystone resources: nut-cracking for oil palm nuts, and pestle-pounding for oil palm pith. Both resources are heavily mechanically protected and cannot be exploited without the use of tools (Yamkoshi 1998). Similarly, Moura and Lee (2004) have argued that “energy bottlenecks” create contexts for capuchins (*Cebus apella*) to derive benefits from tool technology. In the Caatinga dry forest of northeastern Brazil, capuchins have commonly been observed to use tools, and do so during the extended dry season
of this region. During such times, preferred foods are not available; resources that are available without tool use are not sufficient for nutritional requirements, and the capuchins forage terrestrially for tubers. Several tools and tool-facilitated behaviors have been found in four habituated capuchin groups foraging in these areas, including digging for tubers with stones, cracking open seeds and branches with stones; breaking tubers with stones; and using stones as hammers in combination with wooden anvils to crack seeds. These monkeys consume 41 plant foods; tool use increases the use of at least three of these plant species. The researchers argue that Cebus foraging for embedded resources in habitats that experience “energy bottlenecks” is facilitated by innovative tool use (Moura and Lee 2004). A similar argument is made for Cebus apella in another dry region of Brazil, where capuchins commonly use hammer and anvil technology to crack nuts during the dry season when preferred resources are scarce (Fragaszy et al. 2004).

Such observations suggests that foraging-related tool use serves a critical function (sensu Rosenberger 1992; Kinzey 1978); i.e., that, regardless of phenotypic variation (e.g., differences in cultural traditions), these behaviors were initially selected for during critical periods when other, more preferred foods are not available. This may have implications for understanding tool use in early human ancestors (Teaford and Ungar 2000) and can be understood in light of the recent work of Robinson and Wilson (1998), who have pointed that some resources are intrinsically easy to use and are widely preferred, while others require specialized features (in this case, tool use) on the part of the consumer. This allows consumers to exploit nonpreferred resources without compromising their ability to use preferred resources.

11.3 Primate feeding adaptations to a general problem: fiber and plant defense chemicals

- “The plant world is not colored green; it is colored morphine, caffeine, tannin, phenol, terpene, canavanine, latex, phytohaemagglutinin, oxalic acid, saponin, L-dopa, etc. We now hunger for the details . . .” (Janzen 1978 p 73).

The fact that primate evolutionary history is marked by plant consumption is not insignificant from a chemical perspective. Angiosperms have not been passive recipients of animal herbivory over the last 65 Myr; they have instead responded to predation by arming themselves with an extraordinary diversity of defensive chemicals. To date, approximately 12,000 such chemical defenses have
been identified, a number viewed to represent only a fraction of extant plant chemical diversity (Levin 1971, 1976; Freeland and Janzen 1974; McKey 1974; McKey et al. 1978; Rosenthal and Janzen 1979; Gartlan et al. 1980; Harborne 1991). In chemical defense, organic substances are accumulated in plant tissues in such a way that if they are consumed or tasted by an animal, feeding is deterred (Harborne 1991). Such substances may be bitter, have an unappealing odor, be poisonous, or have an antinutritional impact. These compounds are typically categorized into two broad categories: (1) digestion inhibitors, which interfere with the efficiency with which nutrients are obtained by the animal and (2) true toxins, which are harmful to the animal in that they interfere with normal physiology and may result in death (Feeny 1976; Waterman and Kool 1994).

Because most chemical defenses are not directly involved in the primary processes of plant growth and reproduction, they are often collectively referred to as “secondary” compounds or secondary metabolites. However, some authors (Cork and Foley 1991) have argued that despite their roles in primary plant processes, complex structural carbohydrates (fiber) of plant cell walls should be considered defensive chemicals because of their antifeedant/antinutritional effects on consumers. Carbohydrates take the form of monosaccharide sugars, disaccharides, and polysaccharides (NRC 2003). The nonstarch polysaccharides are the fiber components of plant cell walls and can be further divided into the soluble nonstarch carbohydrates (soluble fiber) and the insoluble nonstarch polysaccharides (insoluble fiber). The insoluble, nonstarch polysaccharides comprise the structural components of the plant cell walls (hence, “structural polysaccharides”) and include hemicellulose, cellulose, and lignin. Cellulose is the most abundantly distributed carbohydrate in the world (Sharon 1980) and represents a large proportion of the available energy content of plant foods (Blaxter 1962; Alexander 1993), although no vertebrate has the cellulose-digesting enzyme (cellulase) for breaking down this carbohydrate. The structural carbohydrates must instead be broken down in the gut with the assistance of fungi, protozoans or, most commonly, bacterial symbionts in a process known as fermentation (Lambert 1998). Like cellulose, hemicellulose generally cannot be digested enzymatically but instead is broken down by fermentation (Milton and Demment 1988; NRC 2003). Lignin is completely unavailable to a primate consumer and is refractory to both endogenous enzymes and bacterial fermentation.

Primates have thus evolved in a chemically inhospitable environment and cannot simply consume any plant that they encounter. Indeed, there is ample evidence that primates are extremely selective feeders and consume only a fraction of the plant species available to them in a habitat (Oates 1977, 1987; Glander 1978; Milton 1984). Selective feeding behavior is also exhibited in terms of the times of day a primate will consume a given plant species (quantity/quality of
toxins and antifeedants in a plant ebb and flow throughout the day), as well as which portions of the plant (e.g., leaf tips or petioles, but not the entire leaf), and the total quantity of a particular plant food that is consumed (Glander 1978; Struhsaker 1978; Oates 1987).

In addition to behavioral adaptations, like all herbivorous animals primates have also evolved an array of anatomical and physiological solutions for dealing with plant defenses. Alternative solutions are well illustrated by the extant Cercopithecoida (including Cercopithecinae and Colobinae) and the African Hominoidea. Here, I suggest that the ways in which these taxa ferment fiber and detoxify chemical defenses may have important implications for their density and diversity, both in the Miocene and today.

11.4 Making sense of dietary differences: what monkeys eat, but apes do not

The relationship between body mass and basal metabolism is a negatively allometric one. Smaller mammals thus require relatively more energy to maintain endothermic homeothermy, and larger mammals relatively less (Kleiber 1961; Bell 1971; Jarman 1974; Parra 1978; Gaulin 1979; Schmidt-Nielsen 1984, 1997; Harvey et al. 1987; Martin 1990). This negative relationship can have important implications for diet, with smaller-bodied mammals generally expected to consume a higher-quality diet than larger-bodied mammals (Bell 1971; Jarman 1974; Gaulin 1979).

Yet, contrary to what might be predicted by the so-called “Jarman/Bell principle,” wherever the diets of African apes have been studied in comparison to monkeys, apes are invariably called dietary specialists relying on high-quality foods (relatively low in fiber and toxins; e.g., ripe fruit) in contrast to the smaller-bodied cercopithecoids whose diets often include high levels of fiber and toxins (e.g., leaves, seeds, bark). For example, in Kibale National Park, Uganda, cercopithecines maintain a diverse diet at all times, whereas chimpanzees confine their diet almost exclusively to ripe fruit (Lambert 1997, 2002a, b, 2005; Wrangham et al. 1998). When ripe fruit is scarce, chimpanzees either range further to procure ripe fruit or fall back on the pith of terrestrial herbaceous vegetation. Wrangham et al. (1998) have found that for every month studied over an 11-month period, sympatric chimpanzees had more ripe fruit in their diets than Cercopithecus ascanius, Cercopithecus mitis, and Lophocebus albigena and do not consume any more fiber in the diet than do these much smaller monkeys. Moreover, chimpanzees consume significantly less digestion inhibitors and toxins...
(e.g., condensed tannins, monoterpenoids, and triterpenoids) in their annual diet than cercopithecines.

This pattern holds for other African apes as well. Bonobos (Pan paniscus) focus most of their feeding time on ripe fruits (White 1998). On the basis of the research on mountain gorillas (Gorilla gorilla beringei), gorillas in general were long thought to be folivorous; it is now evident that this subspecies is only folivorous because of a complete lack of succulent fruit in their montaine habitat. Indeed, Nishihara (1999) has found that western lowland gorillas (Gorilla gorilla gorilla) spend 63% of their feeding time on fruit. And in a recent review that included data from all long-term western lowland gorilla sites, Rogers et al. (2004) have found that these apes will maintain fruit in their diet throughout the year; the authors indeed call G. g. gorilla “fruit pursuers, with strong preferences for particular and often rare fruit species, for which they will incur significant foraging costs” (p 175). Lowland gorillas are also highly selective when consuming vegetative plant foods and chose plant parts that contain the least fiber and tannin and most sugar (Remis et al. 2001).

While African apes focus on fruit, cercopithecines in general are noted for their eclectic diet and feeding flexibility (Rudran 1978; Struhsaker 1978; Gautier-Hion 1988; Beeson 1989; Richard et al. 1989; Maisels 1993; Chapman et al. 2002). By the age of one year, Papio cynocephalus in Kenya, for example, have already ingested over 200 food types (Altmann 1998). Foods consumed by various cercopithecines can include those that are both readily digested and also those that are high in structural components; they maintain a higher percentage of nonfruit plant parts regardless of fruit abundance (Rudran 1978; Struhsaker 1978; Cords 1986; Kaplin et al. 1998; Lambert 2002b). In addition, as noted above, cercopithecines have long been noted for their capacity to consume plant foods with greater defensive chemical loads than those consumed by sympatric apes (Waser 1977; Andrews 1981; Conklin-Brittain et al. 1998; Wrangham et al. 1998; Lambert 2000, 2001).

Relative to cercopithecines, colobines tend to be more restricted in terms of food types. These monkeys generally do not consume arthropods for dietary protein, and while some colobine species consume unripe fruit pulp, an alkaline stomach pH (and the potential for acidosis) precludes colobines from consuming high levels of ripe fruit (Kay and Davies 1994). However, colobines are capable of dealing with plant foods high in both fiber and secondary metabolites, and their diet is dominated by seeds and leaves not consumed by other primates.

The dietary differences among monkeys and great apes of the Old World can be explained by alternative fermentation and detoxification systems in Cercopithecoidea and African apes.
11.4.1 Fiber and fermentation

The degree to which the nonlignin structural carbohydrates of plant cell walls can be used as a source of energy depends in part on the length of time that these components are retained in the fermenting chamber(s) of the gastrointestinal tract. The two regions of the primate gastrointestinal tract that have undergone the most specialization for carbohydrate fermentation are the stomach and the large intestine (Chivers and Hladik 1980). In forestomach fermenting primates, the primary fermentation chamber is a modified stomach. Members of the subfamily Colobinae are unique among primates in having this fermentation system (Kay and Davies 1994). Their specialized, complex stomach is divided into four chambers; diverse cellulolytic microorganisms (esp. bacteria) are harbored in the first of these chambers (the forestomach). Caecocolic fermenters are those animals that have an enlarged caecum or colon as the primary fermentation chamber. Enlarged caeco-colic regions in primates are common and found in some species of prosimians, New World monkeys, cercopithecines, and hominoids. Cercopithecines and hominoids in particular have well-developed colons (Hill 1958; Milton 1993; Lambert 1998).

Longer digestive retention times result in higher levels of fermentation and many plant parts require considerable fermentation before they are useful as an energy source (Milton 1981, 1984, 1986, 1993; Lambert 1998; Remis 2000). It is typically argued that “both the total capacity of the digestive tract and the capacity available for microbial fermentation are almost directly proportional to body mass” and that only an increase in body size would allow longer retention times (Cork and Foley 1991 p 139; Kay and Davies 1994). As such, it is presumed that smaller primates will have (both absolutely and relatively) faster food passage rates than larger species, which limit their capacity to ferment fibrous plant components (Parra 1978; Kay 1985; Cork and Foley 1991; Kay and Davies 1994; Van Soest 1996).

Yet, both cercopithecoid subfamilies have relatively longer digestive retention times than the much larger African apes (Lambert 2002c). This is not unexpected among the Colobinae, who have specialized stomachs for consuming a diet high in fiber. On average, their mean digestive retention times range from roughly 40 to 60 h (Lambert 1998; Caton 1999). Neither cercopithecines nor hominoids exhibit such derived anatomy, their simple-stomached gastrointestinal anatomy is very similar. Yet, in an analysis regressing digestive retention times as a function of body size, Lambert (1998) found that the cercopithecines in the analysis were significantly further above the regression line than any other primate taxon. Indeed, despite being on average an order of magnitude smaller than African apes, all tested cercopithecines to date exhibit mean digestive retention times
averaging 31 h (P. troglodytes: 31.5–48 h; G. gorilla: 36.5–61.9 h [Lambert 1998, 2002c; Milton and Demment 1988; Remis 2000]). These digestive results have important implications for understanding how monkeys can consume either similar or greater levels of fiber than larger-bodied apes (Lambert 2002c).

11.4.2 Chemical defenses and detoxification

There are two primary mechanisms for detoxifying plant toxins: (1) microbial activity in a specialized stomach and (2) microsomal enzymes (Freeland and Janzen 1974). Colobines rely on their microbial community in their derived stomach, while cercopithecines and apes must rely on microsomal enzymes. While we know extremely little about such microbial mechanisms in colobines, the potential of the specialized stomach with a diverse and dense microbial community to act as a detoxification chamber has been demonstrated in ruminating ungulates (Keeler et al. 1978; Waterman and Kool 1994).

Mammals without the advantage of forestomach microbial activity rely heavily on microsomal enzymes to detoxify plant toxins (Freeland and Janzen 1974). Microsomal enzymes are located in the endoplasmic reticula of hepatocytes and are activated in the presence of toxins, usually after digesta has left the stomach and entered the intestines. All noncolobine primates rely on this system.

Being smaller in body size can be advantageous for enzyme production since rates of enzymatic activity scale negatively with mammal body size (Walker 1978; Freeland 1991). Freeland (1991) thus suggests that smaller mammals are at an advantage for detoxifying plant secondary metabolites and that the larger the mammal, the greater the preference for foods with low amounts of toxic plant metabolites. These assertions rest largely on results by Walker (1978) who has expressed rates of enzyme activity in rats as a function of liver mass relative to body mass. On average, cercopithecines are smaller than colobines and apes, which may facilitate their consumption of chemically defended plants not available to apes and only available to colobines as a consequence of their specialized stomachs.

Thus, cercopithecoids are at an advantage in their capacity to deal with fiber and plant defenses. Colobines thoroughly digest their high-fiber foods via a specialized stomach and tolerate potentially high levels of plant toxins via a diverse microbial community in this stomach. Cercopithecines, on the other hand, have a simple stomach and instead extract nutrients out of the diet via extensive digestive retention. Staying small relative to apes, cercopithecines are able to maintain high levels of enzyme production for dealing with toxins. Incidentally, staying small also keeps absolute requirements for food low and also facilitates a
faster life history strategy with higher reproductive output than apes. African apes, too, have a simple stomach, but have the digestive retention times expected for their body mass (Milton and Demment 1987; Lambert 1997, 2002c; Remis 2000). These larger-bodied primates avoid exposure to toxins and maintain higher-quality dietary components throughout the year, regardless of their scarcity; in the case of chimpanzees, this is facilitated by tool use. Like other African apes, early hominins probably also generally avoided toxins and high fiber, but then eventually evolved myriad cultural adaptations for breaking down fiber and toxins, including soaking and cooking foods (Aiello and Wheeler 1995; Johns 1996; Milton 1999; Wrangham et al. 1999).

11.4.3 Ecological and evolutionary implications

These digestive and feeding adaptations maybe tied to both extant and extinct patterns of species density and diversity. Remaining fairly small keeps cercopithecoid absolute requirements for foods low relative to apes. Moreover, the carrying capacity for monkeys in a given habitat is essentially higher: a function of the fact that more items are food for cercopithecoids. These two factors may account for striking differences in monkey and ape density and biomass in Africa. For example, cercopithecoid density in Kibale National Park, Uganda, is 420 kg/km², while chimpanzee density is 3 kg/km²; the biomass differences are 2,611 and 85 kg/km², respectively (Chapman and Lambert 2000). This pattern holds elsewhere in Africa. In Tai Forest, for example, the cercopithecoid biomass compared to that of chimpanzees is 951.7 versus 58.3 kg/km² (Chapman et al. 1999). In Budongo, Uganda, the biomass of cercopithecines is 354, and chimpanzees 89 kg/km², and in the Lope Reserve, Gabon, where both P. troglodytes and G. gorilla are found, the monkey biomass is 251.4 kg/km², and ape biomass is 65.8 kg/km² (Plumptre and Reynolds 1995; Chapman et al. 1999).

Extant cercopithecoids are also greatly more speciose than apes. As has been noted by a number of authors, this pattern of diversity is a relatively recent one and has changed dramatically since the Miocene. During that epoch (23–5 Ma), hominoid species richness was at its maximum. In their species diversity and range of ecological adaptations and body sizes, Miocene apes paralleled the diversity of modern Cercopithecidea (Andrews 1981; Kelley 1992). Since the Miocene, however, hominoid species diversity has steadily decreased, from a maximum of four to six sympatric species in the earlier millennia of the epoch, to no more than two species living in a single habitat in the later Miocene and present (Andrews 1981). Conversely, while monkey diversity was relatively low in
the Early Miocene, it has increased steadily, eventually achieving levels comparable to extant patterns.

Several explanations have been proposed to account for this shift in hominoid–cercopithecoid diversity (Napier 1970; Ripley 1979; Andrews 1981; Temerin and Cant 1983). Each explanation is based on the concomitant assumptions that there was competition for increasingly scarce food resources between Miocene monkeys and apes and that this competition resulted in an ape adaptive pattern and a monkey adaptive pattern for foraging and feeding. Early African monkeys and apes are viewed to have been in competition for increasingly scarce fruiting resources during a time of extensive environmental change (cooler, dryer) in Miocene Africa (Andrews and Van Couvering 1975). Napier (1970) argued that, in the face of increasingly patchy fruit availability in a changing East African habitat, the divergence of the cercopithecoids was directly related to their ability, via specialized bilophodont molars, to consume leaves. Similarly, Ripley (1979) argues that increasing East African forest seasonality played an important role in Old World monkey–ape divergence, although rather than dentition, she places emphasis on locomotor differences. Andrews (1981) refined Napier’s (1970) hypothesis on cercopithecoid–hominoid divergence and suggested that monkeys diverged from apes as a result of their being able to exploit unripe fruits. Like Napier’s scheme, his support stems largely from dental evidence.

Temerin and Cant (1983) proposed a model of ecological energetics to explain the differences in monkeys and apes. In this scheme, the authors argue that “... when consuming the same category of food items (ripe fruit, leaf shoots, young leaves, etc.) apes exploit more widely distributed and/or smaller patches on average than do OWM” (Temerin and Cant 1983 p 343). Their fundamental premise is that in the same environment, with access to the same resources, monkeys emphasize energy gain, while apes decrease energy expenditure because of their more efficient, specialized locomotion. They argue that in a habitat with increasingly rare fruit resources, Miocene monkeys did not move greater distances to maintain a heavily frugivorous diet; rather, they shifted to other food types. Conversely, apes did not shift their dietary proportions but instead traveled longer distances to exploit increasingly rare fruit; this was facilitated by their specialized postcrania and efficient locomotion.

Several predictions have been made regarding Old World monkey and ape ecological distinctions and the shift in hominoid–cercopithecoid diversity from the Miocene to the present (Napier 1970; Ripley 1979; Andrews 1981; Temerin and Cant 1983). For example, Temerin and Cant (1983) predicted that, in a given environment, monkeys will eat more leaves and unripe fruit than apes, and that monkeys will have higher assimilation efficiencies than apes on diets containing plant fiber. Similarly, Andrews (1981 p 49) has argued:
a critical distinction between apes and monkeys is the greater tolerance of the latter for a variety of plant secondary compounds, including tannins and alkaloids. Conversely, hominoid species avoid food with high levels of toxicity, and in particular they avoid unripe fruit that is eaten by monkeys. In developing a tolerance for tannins and other secondary compounds, therefore, the cercopithecines are able to tolerate less ripe fruit than the hominoids and thus gain access to fruiting trees before the hominoids.

Clearly, these predictions have been borne out in the last 24 years of primate field research. Yet, to date, there have been no explicit hypotheses regarding what these “assimilation efficiencies” and “tolerances” are, or how they work. I suggest that they are specialized features of digestive and detoxification systems, including either long retention times/relatively high rates of microsomal enzyme production in the case of cercopithecines, or specialized stomach and microbial action in the case of the colobines.

On the basis of the variation in shearing quotients (SQ) in fossil apes, Ungar and Kay (1995) argue that the dietary breadth of Miocene hominoids was much greater than it is today, with fossil ape species occupying a variety of trophic niches, including folivory, soft-object frugivory, and hard-object feeding (seeds). That Miocene apes ranged in size from the small Micropithecus clarki at 3.5 kg to Afropithecus and Proconsul major weighing in at about 50 kg (Fleagle 1988) has been used in support of arguments regarding dietary range in Miocene apes. However, my proposition deflates the theoretical importance of body mass in driving the Miocene monkey/ape divergence. While it is commonly held that large bodied primates have a greater capacity for fermentation and longer retention times (Gaulin 1979; Fleagle 1988), it appears that, in fact, cercopithecines successfully evolved this capacity without necessarily evolving large body size. Smaller mammals do have relatively greater metabolic needs than larger ones. However, rather than requiring a high-quality diet that is digested and processed quickly in order to maintain a high rate of incoming food, cercopithecines simply make more out of a lower-quality diet via detoxification and fermentation. Indeed, I suggest that from a dietary perspective, smaller, “monkey-sized” apes of the Early Miocene were not necessarily analogs of later similarly sized monkeys. Although Miocene apes may have had the dental features correlated with folivory and seed-eating (i.e., high SQ), they may have ultimately been out competed by cercopithecoids, because monkeys had both dietary and digestive features that enabled them to deal with a broader diet (i.e., colobines had a sacculated stomach and microbial detoxification, and cercopithecines long retention times and high rates of enzyme production), and were simply better at extracting nutrients/resources out of a variety of food types. Thus, cercopithecoids of the Miocene
and thereafter were better equipped to deal with the variety of niches originally occupied by the dental apes. Apes of the Middle to Late Miocene may have been essentially outcompeted (digestively) in these other niches and, as a result, became better frugivores, resulting in the highly selective fruit-feeding behavior we see today.

In short, I suggest that at some point early in their radiation, cercopithecinines increased the breadth of their dietary niche and did so by remaining small and evolving long retention times. As the cercopithecoids diverged, colobines took off on a different evolutionary trajectory and opted for a previously unfilled primate dietary niche (specialized arboreal folivore) by evolving a specialized stomach. Relative to apes, cercopithecoids have a faster life history strategy with much greater reproductive output; a consequence of this is that smaller primates tend to be more speciose and evolve more quickly, which too may explain their greater speciosity (Cowlishaw and Dunbar 2000). Hominoids opted for a specialized strategy and are essentially “trophically restricted” (sensu Ungar and Kay 1995) relative to the smaller cercopithecoids, while hominins pursued a trajectory of increasing dietary quality by both increasingly including vertebrate meat into the diet and by cultural adaptations such as cooking (Milton 1999; Wrangham et al. 1999; Stanford and Bunn 2001).

References


Freeland WJ (1991) Plant secondary metabolites: biochemical coevolution with...


Lambert JE (1997) Digestive strategies, fruit processing, and seed dispersal in the chimpanzees (Pan troglodytes) and redtail monkeys (Cercopithecus ascanius) of Kibale National Park, Uganda. Ph.D. dissertation, University of Illinois, Champaign-Urbana


Parker ST, Gibson KR (1977) Object manipulation, tool use and sensorimotor intelligence


Rudran R (1978) Socioecology of the blue monkeys of the Kibale Forest, Uganda. Smithsonian contributions to zoology, vol 249. Smithsonian Institution Press, Washington DC


cultural variation in chimpanzees. Behaviour 138: 1481–1516
12 Great Ape Social Systems

Angela Meder

Abstract

For decades, the social systems of the great ape species were described as being fundamentally different, but long-term field studies have questioned this idea. Although orangutans seem to be less social than the African apes, they have the ability to socialize and spend much time in contact with conspecifics if the ecological conditions permit. Although the social behavior of chimpanzees and bonobos seems very different, their social structure is similar. The social systems of all great apes have common traits. They are characterized by weak ties, female transfer is common, they have a tendency toward fission–fusion grouping and a complex social network, females lack sharply defined dominance relations, and intrasexual bonds among nonkin can be relatively strong. There is remarkable intraspecific variability in social organization and structure; differences in ecological conditions seem to be very important in determining this. Especially frugivory requires a mobile and flexible social system. Compared to the great apes, humans show an even greater variability.

12.1 Introduction

Great apes share a common ancestor, they live in similar habitats (at least some populations), they are rather large, and they have slow life histories. Despite these similarities, they developed different social structures. What are the similarities? What are the differences and why did they develop? What social structure might the common ancestor have had? What does this mean for Homo, who emerged from within the great apes?

These questions have occupied primatologists for decades. Many hypotheses and models were developed and had to be abandoned because the growing evidence from field research did not support them. Some questions cannot be answered satisfactorily yet, but we have a much better idea now of the social systems of our closest relatives. This knowledge is summarized here. Ecology and life history are also briefly portrayed for each species because they are key factors for the understanding of group structures and dynamics.
12.2 Primate social systems

Animals may live in groups to increase their survival and reproductive success, but these groups have different structures for each species. To explain social systems, Kappeler and van Schaik (2002) distinguish three components: social organization (group size and organization), mating system, and social structure (relationships between group members). These three components are used here to clarify the social structures of the ape species.

Dunbar (1988) lists four reasons why primates should associate (stay and move) with conspecifics: better protection against predators, defense of resources, foraging efficiency, and improved caregiving opportunities. But sociality also has costs—especially higher intraspecific competition. Competition means usually competition for food (Kummer 1971; Wrangham 1979). Large groups deplete food more quickly and have to travel further in a day; this may influence female grouping patterns and restrict travel and sociality of females with infants (Sommer and Reichard 2000).

Folivores do not need to travel far each day and can live in relatively large groups without high levels of competition. The distance a group of a frugivore species has to travel is related to group size, therefore food distribution limits the number of group members (at least during feeding). Therefore, frugivores should either live in small groups or in fluctuating grouping patterns.

According to Dunbar (1988), several other factors may also influence primate group structure—at least theoretically. An important factor in this respect is the risk of male infanticide. Females with long lactation (like great apes) face an increased risk of infanticide, particularly when a reproductively capable male comes into a position of top dominance (van Schaik 2000). Usually male infanticide is seen, especially in one-male groups, as a strategy to increase reproductive success: the female will become receptive again sooner and the bond between the partners will be strengthened.

Regarding reproductive systems, the ape species have traditionally been classified as follows (Sommer and Reichard 2000):

- one male–many females (*polygyny*): gorilla
- one male–one female (*monogamy*): gibbons
- many males–many females (*polygynandry*): chimpanzees
- many males–one female (*polyandry*): not generally present in apes

As in this list, certain social systems have often been attributed to certain mating systems, but this is much too simple—not only in apes but also in
primates in general. A social unit has its own history; a group may be polygynous/polyandrous, for example, and temporarily this may mean monogamy, but one individual may have several successive partners during its lifetime. Moreover, primates, and especially the great apes, show considerable flexibility and variability in their social behavior.

Finally, grouping levels may form a hierarchy. In great apes, the fission–fusion society is such a multilevel social system. Fission–fusion means that groups with different size and different composition may form for limited periods of time but that above this lowest grouping level, there is a higher, stable social unit.

12.3 The “lesser” apes

Although this chapter concentrates on the family Hominidae, the sister family Hylobatidae will be briefly discussed. Gibbons are much smaller than the great apes: they weigh about 5–11 kg, with the siamang being the largest species. They live mostly in evergreen rain forest, some populations in semideciduous and mixed evergreen forest are strictly arboreal and are primarily frugivorous.

For a long time, gibbons were regarded as one of the few monogamous primate taxa, but recently this concept has been questioned. Change or replacement of a partner occurs in many species; in Khao Yai (Thailand), where the population density is very high, there is intense competition for females. The majority of adults copulate or attempt to copulate with partners other than their mates. Social polyandry is not uncommon, and male immigration into existing monogamous groups is observed. Extensive overlap of territories and a wide range of affiliative interactions with neighbors have been described in several gibbon species (Sommer and Reichard 2000). They show a high interspecies variation in social structure, but even if Brockelman et al. (1998) question it, their basic social structure seems to be the nuclear family unit.

Traditionally, the male–female relationship was regarded as pair-bonded, but it seems that gibbons rather live in small groups, most probably because selective pressures do not permit large groups. At least in some species, these social units are not really two-adult groups, although it is the norm. Sommer and Reichard (2000) suggest that male gibbons associate with females in order to provide a constraint to roving males and thus protect their offspring from infanticide.

Fuentes (2000) hypothesized that gibbons may live in a supragroup organization—a variable community that consists of a set of relatively cohesive bisexual clusters that may have evolved from an original fission–fusion community organization. This model, however, is not supported by evidence from the field.
12.4 Orangutans (genus *Pongo*)

Bornean and Sumatran orangutans are today usually classed as two distinct species (Table 12.1). There seem to be characteristic differences between the two species in behavior and social organization.

12.4.1 Ecology

Orangutans live in rain forest habitats with a high tree species diversity from sea-level swamp forests to mountain slopes (Galdikas 1988; Knott 1999). They prefer the lowland forest up to about 1,000 m, usually close to streams and rivers as well as swamp. Galdikas (1988) lists 317 food types for Tanjung Puting. The orangutan diet consists of 61% fruit; figs are their staple food. They eat meat (small mammals) on rare occasions (Rijksen and Meijaard 1999). Part of their habitat is characterized by extreme fluctuations in the type and quantity of fruit available, and this results in dramatic differences in nutritional intake and body weight (Knott 1999).

Orangutans require a large area that typically includes various habitats. Resident females have stable, overlapping home ranges with an area of 0.5–6 km² on Borneo and 1–3 km² on Sumatra. The ranges are not actively defended and overlap considerably. The size of a resident adult male’s home range is 1–8 km² on Borneo and 6–10 km² on Sumatra. It is neither exclusive nor stable (van Schaik and van Hooff 1996; Rijksen and Meijaard 1999; Singleton and van Schaik 2002). Daily travel distances vary from 90 m to 3.1 km (mean: 0.79 km) in Tanjung Puting.

Orangutans are the largest arboreal animals. Females move almost exclusively at between 10- and 25-m height in the trees. In Tanjung Puting, males sometimes walk on the ground, where they also forage and rest, but not in Sumatra, where tigers occur (Sugardjito et al. 1987; Galdikas 1988).

12.4.2 Life histories and dispersal

Orangutans may be nursed until 6 years old, longer than any other ape species. The mean interbirth interval is 8 years (Leighton et al. 1995). Females may conceive from about 11 years (Knott 1999). Fully adult (flanged) Bornean males weigh 86.2 kg on average, females 38.7 kg; a single Sumatran male weighed 86.2 kg and Sumatran females on average 38.3 kg (Markham and Groves 1990). Apparently mature but unflanged males weigh about 60–70 kg. In captivity,
## Table 12.1
Great ape species and subspecies and their distribution

<table>
<thead>
<tr>
<th>Species/Subspecies</th>
<th>Distribution</th>
<th>Study sites mentioned here</th>
<th>Habitat</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Pongo abelii</em> Sumatran orangutan</td>
<td>Sumatra</td>
<td>Gunung Leuser (Ketambe, Suaq Balimbing)</td>
<td>Rain forest, swamp forest</td>
</tr>
<tr>
<td><em>Pongo pygmaeus</em> Bornean orangutan</td>
<td>Borneo</td>
<td>Tanjung Puting, Kutai</td>
<td>Rain forest, swamp forest</td>
</tr>
<tr>
<td><em>Gorilla beringei</em>—eastern gorilla</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Gorilla beringei beringei</em>—mountain gorilla</td>
<td>Virunga Volcanoes, Bwindi (Uganda)</td>
<td>Virunga Volcanoes, Bwindi</td>
<td>Montane forest, occasionally grassland</td>
</tr>
<tr>
<td><em>Gorilla beringei graueri</em>—eastern lowland gorilla, Grauer’s gorilla</td>
<td>Eastern D. R. Congo</td>
<td>Kahuzi-Biega (mountains), Utu, Masisi (extinct)</td>
<td>Lowland rain forest, montane forest, occasionally swamps</td>
</tr>
<tr>
<td><em>Gorilla gorilla</em>—western gorilla</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Gorilla gorilla diehli</em> Cross River gorilla</td>
<td>Cross River area (Nigeria/Cameroon)</td>
<td>Cross River National Park, Afi, Takamanda</td>
<td>Lowland rain forest, montane forest</td>
</tr>
<tr>
<td><em>Gorilla gorilla gorilla</em> Western lowland gorilla</td>
<td>From Cameroon to Cabinda in the west to the Central African Republic and the Congo Republic in the east</td>
<td>Lopé, Nouabalé-Ndoki (Mbeli Bai), Dzanga-Sangha (Bai Hokou), Lossi, Odzala, Río Muni, Lokoué Bai</td>
<td>Lowland rain forest, occasionally grassland and swamps</td>
</tr>
<tr>
<td><em>Pan troglodytes</em>—chimpanzee</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Pan troglodytes schweinfurthii</em> Eastern chimpanzee</td>
<td>D. R. Congo, north of the Congo River from its junction with the Oubangui into Sudan and the Ituri region to Lakes Edward and Albert</td>
<td></td>
<td>Rain forest, woodland</td>
</tr>
</tbody>
</table>
Pan troglodytes marungensis
Southeastern D. R. Congo to the Rutshuru district, Uganda, Rwanda, Burundi, Tanzania
Gombe, Mahale, Bwindi, Budongo, Kibale, Semliki, Kahuzi-Biega
Woodland, gallery forest, semi-deciduous forest, lowland and montane rain forest

Pan troglodytes troglodytes
Central chimpanzee
From the Congo River mouth to the Congo Republic and the Central African Republic to Cameroon, north to the Sanaga River
Lossi, Lopé, Río Muni, Ndoki
Rain forest

Pan troglodytes vellerosus
East Nigeria-West Cameroon chimpanzee
Nigeria and Cameroon to the Sanaga River
Gashaka
Lowland forest, montane forest, woodland

Pan troglodytes verus
West African chimpanzee
West Africa to Togo in the east
Tai, Assirik, Bossou
Savanna, deciduous forest, rain forest

Pan paniscus
Bonobo
D. R. Congo, south of the Congo River
Lomako, Wamba, Lukuru
Rain forest, swamp forest, occasionally grassland, dry forest

Table 12.1 (continued)

<table>
<thead>
<tr>
<th>Species/Subspecies</th>
<th>Distribution</th>
<th>Study sites mentioned here</th>
<th>Habitat</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Pan troglodytes</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>marungensis</em></td>
<td>Southeastern D. R. Congo to the Rutshuru district, Uganda, Rwanda, Burundi,</td>
<td>Gombe, Mahale, Bwindi, Budongo, Kibale, Semliki, Kahuzi-Biega</td>
<td>Woodland, gallery forest, semi-</td>
</tr>
<tr>
<td></td>
<td>Tanzania</td>
<td></td>
<td>deciduous forest, lowland and</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>montane rain forest</td>
</tr>
<tr>
<td><em>Pan troglodytes</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>troglodytes</em></td>
<td>From the Congo River mouth to the Congo Republic and the Central African</td>
<td>Lossi, Lopé, Río Muni, Ndoki</td>
<td>Rain forest</td>
</tr>
<tr>
<td><em>Central chimpanzee</em></td>
<td>Republic to Cameroon, north to the Sanaga River</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Pan troglodytes</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>vellerosus</em></td>
<td>Nigeria and Cameroon to the Sanaga River</td>
<td>Gashaka</td>
<td>Lowland forest, montane</td>
</tr>
<tr>
<td>*East Nigeria-West</td>
<td></td>
<td></td>
<td>forest, woodland</td>
</tr>
<tr>
<td>Cameroon chimpanzee*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Pan troglodytes</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>verus</em></td>
<td>West Africa to Togo in the east</td>
<td>Tai, Assirik, Bossou</td>
<td>Savanna, deciduous forest, rain</td>
</tr>
<tr>
<td><em>West African chimpanzee</em></td>
<td></td>
<td></td>
<td>forest</td>
</tr>
<tr>
<td><em>Pan paniscus</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Bonobo</em></td>
<td>D. R. Congo, south of the Congo River</td>
<td>Lomako, Wamba, Lukuru</td>
<td>Rain forest, swamp forest, occasionally grassland, dry forest</td>
</tr>
</tbody>
</table>

Taxonomy according to Groves (2001); Groves (2005); the Bornean orangutans’ taxonomy is still in discussion (Kanthaswamy and Smith 2002).
orangutans have reached as much as 58 years of age, whereas in the wild their maximum age is estimated at 45 years (Leighton et al. 1995).

On reaching adulthood, males develop typical secondary sexual characteristics: flanged cheeks and a big throat sac which enables them to produce loud calls (“flanged” males). Some males show an extended subadult appearance (“unflanged” males), in some cases until the age of 30. This parallel existence of two morphs of mature males is called bimaturism (Utami Atmoko and van Hooff 2004). Orangutan male bimaturism is socially influenced; the close presence of a flanged male suppresses the development of secondary sexual characteristics in subordinate males (van Hooff 1995).

Females often stay in ranges that overlap or are close to their natal area (Galdikas 1984; Singleton and van Schaik 2002). Despite these observations, Utami et al. (2002) found that adult males and females are mostly unrelated in Sumatra—this means that both sexes disperse. Males generally seem to disperse more widely (van Hooff 1995; van Schaik and van Hooff 1996).

Regarding their ranging behavior, Rijksen and Meijaard (1999) discern three types of orangutans: residents, who stay in the same area for years; commuters, who appear regularly and have a very large home range; and wanderers, who are seen very infrequently and seem to have no home range at all. There are considerable between-site differences in the percentage of these types. In some sites on Sumatra no residents are found at all. In Ketambe, commuters are the majority of the population (Rijksen and Meijaard 1999; van Schaik 1999).

12.4.3 Size and structure of associations

Both species of orangutans move independently most of the time. Usually, trees do not provide enough food for several adults. It is difficult to get statistical data on the sizes of associations, because all field studies bias toward larger party sizes, since they are found more easily. The mean daily party size is 1.67 independent individuals in Suaq Balimbing, and the largest sustained party size that has been observed so far was 10.67 (van Schaik 1999).

It is sometimes difficult to decide whether orangutan associations can be called social or whether the apes just meet during their search for food and tolerate the presence of each other. According to Galdikas (1984), in Kutai only 6 out of 13 groups could be considered “social”: a consorting adult pair and two female-offspring units. In Tanjung Puting, Galdikas (1995) observed that 94% of the time adolescent females are in contact with others included true social interaction, the rest is nonsocial aggregation. In other age/sex classes, roughly 80%–90% of associations are true social groupings.
The most common grouping is one subadult male and one unreceptive adolescent female (Galdikas 1988). Subadult males and adolescent females spend about 40% of their time in contact with other units in Tanjung Puting, adult males 16.8% and adult females 13.5% (Galdikas 1995). Adult females in Borneo are far more solitary than those in Sumatra; this reflects the difference in average orangutan density (van Schaik 1999). Fully adult males suffer energetically from association and thus are extremely solitary (van Schaik and van Hooff 1996). The largest stable parties in Suaj Balimbing develop when females converge upon a dominant male. Sexual associations seem to last longer there and involve more animals than elsewhere (van Schaik 1999).

12.4.4 Male–female relationships and mating strategies

Contacts of adolescent females with adult males in Tanjung Puting are almost totally restricted to consortships which the female initiates, maintains, and terminates. In general, adult females either avoid contact with males or ignore them, while adolescent females (9–14 years) spend 56% of their contact time exclusively with males. Toward adult males, nonreceptive adolescent and adult females show avoidance or indifference; if they are receptive, however, they are highly motivated to maintain contact with the seemingly indifferent adult males and may respond to their long calls by immediate approach. With subadult males, adolescent females have long-term associations with no copulation, which may be called friendships, as well as short associations, but no consortships. The subadult males are responsible for maintaining contact with nonreceptive adolescent females, and sometimes they travel with them for weeks (Galdikas 1984, 1995).

There are two behavioral tactics in the mating strategy of male orangutans: resident flanged males sit, utter long calls, and wait for females (residents, commuters, and wanderers), while unflanged and nonresident flanged males actively search for females and mate with them, irrespective of their receptivity (Utami Atmoko and van Hooff 2004). Rijksen and Meijaard (1999) assume that these rapes are a sexual strategy of young males because the ascent to a higher social status is associated with long-lasting sexual relationships with several females—and this status may be the prerequisite for the development of the full sexual characteristics (Utami et al. 2002). During consortship, the partners travel closely for days or even weeks, usually with the adult male following (Galdikas 1984; Singleton and van Schaik 2002). Severe mating competition is usual among fully adult orangutan males, but no infanticide attempts have ever been seen so far.

Both mating tactics are successful. A paternity analysis in Ketambe, Sumatra, showed that all but 1 of 11 infants were sired by resident males; out of these
10 infants, 6 were sired by unflanged males (Utami et al. 2002). In general, however, unflanged males seem to have a slightly lower reproductive success (Satkoski et al. 2004).

### 12.4.5 Female–female relationships

Competition for access to limited resources is rare. On Sumatra, where the population density is high, female home ranges overlap considerably and the rates of association between females are high compared to Borneo (Singleton and van Schaik 2002). The longest association Galdikas (1984) observed between two adult females in Tanjung Puting was 3 days and 2 nights. Sometimes females travel together, but few direct interactions are observed. Their behavior toward each other varies from aggression and avoidance to affiliative activities, but only very rarely does a female touch another female affiliatively.

Adolescent females often travel together after encounters. The longest association Galdikas (1995) saw lasted at least 10 days. Although adult females are often intolerant to and may attack adolescent females, the adolescents try to associate with them.

### 12.4.6 Male–male relationships

Flanged males avoid each other. They produce long calls that carry for up to 1 km—especially when they approach another individual, after they have defeated another adult male or before they travel over long distances. When adult males come close to each other, agonistic interactions are observed. They have the highest incidence of disfigurements among orangutans, presumably from injurious fights. They dominate unflanged males but may tolerate them as long as they keep a certain distance (van Schaik and van Hoooff 1996).

### 12.4.7 Discussion: genus Pongo

An older model of orangutan social relationships is that a fully adult male controls a relatively large range that includes the smaller ranges of several females. Calculations proved, however, that a male can maximally defend 0.37–0.57 km², which is less than the size of a single female’s range. Moreover, adult male ranges overlap and mating is promiscuous (van Schaik and van Hoooff 1996).

Orangutans have been regarded as solitary, but they have the ability to socialize and spend much time in contact with conspecifics if the ecological
conditions permit. That they are not antisocial is documented from zoos, where they have been kept in rather large groups (Poole 1987). Restrictions on group size must therefore lie in the environment in the wild.

To associate with conspecifics in feeding parties is costly if the food in one tree is not sufficient for several adults (especially for large males). But social contacts are important, particularly for the development of various behaviors, of social competence and relationships. Females with new offspring therefore form "nursery groups" in which new and weaned offspring can socialize (Galdikas 1995; van Schaik 1999). Infants learn there how to behave correctly toward conspecifics. Even in independent immature animals, social contact is still necessary for further development. According to Galdikas (1995), adolescent females use social contacts to establish their own home range and to form relationships with the individuals in adjacent ranges. Sugardjito et al. (1987) discuss more benefits of grouping.

Orangutans show the greatest degree of sexual dimorphism of all great apes. Such an extreme dimorphism may be a result of female choice or male–male competition. Other species with such sexual dimorphism are usually either living in one-male groups or have a lek-type mating system. Rijksen and Meijaard (1999) suggest that orangutans have a very complex lek-type mating system. Males meet in a "social arena," an area with especially abundant food resources, where the ranges of several residents overlap and where commuters and wanderers visit regularly. In that arena they emit long calls to attract females.

Two models for the description of the orangutan social system are discussed by van Schaik and van Hooff (1996): either a roving male promiscuity system (an estrous female attracts multiple males who compete vehemently for sexual access to her) with no higher-level social unit or a spatially dispersed but socially distinct community organized around one or more large adult males. Both models have weaknesses: well-defined communities do not seem to exist in any orang population and orangutan females do not mate with any available male. They seem to show distinct preferences; individual relationships and bonding are important (Galdikas 1984; van Schaik and van Hooff 1996; Rijksen and Meijaard 1999; van Schaik 1999; Utami et al. 2002). The roving male promiscuity system seems to be closer to the situation of the Bornean orangutans, and the socially distinct community system organized around a dominant male fits better for the Sumatran population (Singleton and van Schaik 2002).

Although orangutan communities comparable to those of chimpanzees do not exist, orangutans obviously have a social network. Residents and commuters in a certain area seem to know each other well, while wanderers are strangers to them. An explanation could be provided by a recent suggestion by Rijksen and Meijaard (1999) and van Schaik (1999) that the individual-based fission–fusion
social system that is characteristic for chimpanzees may also be typical for orangutan societies; the differences are in degree rather than in quality.

12.5 **Gorillas (genus *Gorilla*)**

Eastern and western gorillas are today usually classed as two distinct species (Table 12.1). Long-term studies are available now for both species from several sites.

12.5.1 **Ecology**

Gorillas live in a variety of habitats: lowland rainforest, swamp forest, marshy clearings (bais), and montane forest. The lowest densities of western lowland gorillas are observed in areas without Marantaceae and Zingiberaceae, their preferred food plants, and the highest densities in Marantaceae and swamp forest (Poulsen and Clark 2004; Rogers et al. 2004).

Mountain gorillas mainly feed on green plant parts, whereas lowland gorillas eat a lot of fruit (but still markedly less than chimpanzees and orangutans). Utilization of fruit and of herbaceous vegetation varies seasonally. Western lowland gorillas eat about 250 food items with up to about 100 fruit species, while the mountain gorillas of the Virunga Volcanoes eat only up to 72 different plant species and very few fruits (Tutin and Fernandez 1993; McNeilage 2001; Doran et al. 2002a). Gorillas do not kill vertebrates but eat small animals, mainly insects.

Gorillas do not occupy territories. Their annual home range covers about 8 km² in the Virunga gorillas, about 30–40 km² in eastern, and 11–30 km² in western lowland gorillas. The size depends on the distribution of food sources and group size; home ranges overlap (Tutin 1996; Yamagiwa et al. 1996, 2003a; Remis 1997; McNeilage 2001; Bermejo 2004). In general, gorilla groups move an average of 0.5–2 km a day to forage. Folivorous gorilla populations in montane forest show a shorter daily journey length than more frugivorous populations (Yamagiwa et al. 2003a, b).

Apart from humans, gorillas do not really have “predators.” The only predator is the leopard (Tutin and Fernandez 1991; Fay et al. 1995). They live primarily on the ground, particularly in the Virunga Volcanoes, where females spend only 7% of the time above the ground and males 2% (Doran and McNeilage 1998). Silverback males do not often leave the ground because of their great weight. Gorillas sleep on bare earth or in nests, which they build on the ground or in trees (Poulsen and Clark 2004).
12.5.2 Life histories and dispersal

Gorilla infants are nursed for at least 2–3 years, the interbirth interval is about 4 years. Gorillas grow faster and breed more rapidly than other hominids (Groves and Meder 2001). Females may first reproduce between about 8 and 9 years of age in the wild (Watts 1991; Robbins et al. 2004). Males are fully grown at approximately 15 years of age and are called silverbacks. Young mature males without the secondary sexual characteristics are called blackbacks. Western lowland gorilla males have a mean weight of 149 kg and females 71.5 kg, mountain gorilla males weigh 157 kg and females 97.7 kg, and Grauer’s gorilla males 171 kg and females 80 kg (Meder 1993; Rowe 1996). Adults have a relatively short life expectancy; silverback males, in particular, seem to have a hard life and die young (Groves and Meder 2001). In the wild, they probably reach 40–45 years; in captivity, several individuals have lived more than 50 years.

Reaching adulthood, female gorillas usually leave the group they were born in and join a new unit. They emigrate only if they encounter another male. Often they transfer several times before they settle down in a certain group (Watts 1996; Sicotte 2001; Stokes et al. 2003). In Kahuzi-Biega, the simultaneous transfer of several individuals was observed (Yamagiwa and Kahekwa 2001).

In western gorillas and Grauer’s gorillas male emigration is common, while among mountain gorillas less than 50% of the males emigrate. They either become solitary or join all-male groups in some populations (Watts 2000; Yamagiwa et al. 2003a; Robbins et al. 2004). In western lowland gorillas (Lokoué Bai, Congo), even subadults and juveniles may emigrate (Gatti et al. 2004).

12.5.3 Size and structure of social units

In a gorilla group there is a clear hierarchy. The leading silverback has the highest rank, and adult females are dominant over young animals. In most populations, groups usually consist of one adult male, several females, and their offspring (Magliocca et al. 1999; Yamagiwa et al. 2003a; Gatti et al. 2004). In the Virunga mountains, in Bwindi and Bai Hokou, however, up to 53% of all groups include more than one adult male (Table 12.2; Goldsmith 2003; Kalpers et al. 2003).

As groups contain more females than males, many males are “left over.” Gatti et al. (2004) found that at Lokoué Bai 42% of the silverbacks are unmated; 31% are solitary and 11% live in nonbreeding groups. Solitary males make up 5%–10% of the western lowland gorilla populations. Among eastern gorillas, the percentage seems to be lower: about 3.5% in Kahuzi-Biega and 1.8%–2.7% in mountain gorillas, which may reflect differences in the rate of natal emigration (Parnell 2002).
Gorillas occasionally form all-male groups, usually with one mature male and a few younger males. Such nonreproductive groups have a similar size to breeding groups; they have been observed in western lowland as well as mountain gorillas, although not in Grauer’s gorillas (Robbins et al. 2004). These groups as units may be long-lasting, but their membership changes quite often due to male migrations. They seem to be transition units in both gorilla species (Gatti et al. 2004).

Gorilla groups can have very different histories (Robbins 2001). When the dominant male dies, the group may disperse if no subordinate silverback is there to take over the leadership; if there are two younger silverbacks, the group may split. Parnell (2002) lists five stages in the development of a typical western lowland gorilla group: nascent (one silverback + one female without offspring); infant (one silverback + one or more females with dependent offspring); juvenile (one silverback + one or more females with a range of offspring); mature (one

<table>
<thead>
<tr>
<th>Region</th>
<th>Group size</th>
<th>Sample size</th>
<th>Number of silverbacks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mt. Alen, Río Muni</td>
<td>7.13</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Abumnzok- Añinzok, Río Muni</td>
<td>6.4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Ndoki, Congo</td>
<td>7.3</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Mbeli Bai, Congo</td>
<td>8.4</td>
<td>14</td>
<td>1.1</td>
</tr>
<tr>
<td>Maya Nord, Odzala, Congo</td>
<td>11.2</td>
<td>31</td>
<td>1.0</td>
</tr>
<tr>
<td>Lossi, Congo</td>
<td>17</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Lopé, Gabon</td>
<td>9.3</td>
<td>4</td>
<td>1.25</td>
</tr>
<tr>
<td>Kahuzi-Biega, mountains</td>
<td>14.3</td>
<td>12</td>
<td>1.2</td>
</tr>
<tr>
<td>Kahuzi-Biega, lowland (Utu)</td>
<td>15.6</td>
<td>14</td>
<td>1.1</td>
</tr>
<tr>
<td>Masisi</td>
<td>10.0</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>Virunga volcanoes</td>
<td>15</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Bwindi, Uganda</td>
<td>8.8</td>
<td>28</td>
<td>1.75</td>
</tr>
</tbody>
</table>
silverback + one or more females and a range of offspring from infants to young silverbacks); and senescent (one old silverback, few or no females, only older offspring). Group takeovers by outside males have not been reported so far (Robbins et al. 2004), but in zoos new males have been introduced successfully to established groups in many cases.

The size of gorilla groups is very variable but similar among the subspecies (Table 12.2). An average group contains about nine to ten members (Parnell 2002; Yamagiwa et al. 2003a). The largest group observed so far was Pablo’s group in Rwanda with 60 members in May 2006.

12.5.4 Male–female relationships and mating strategies

In the social system of gorillas, the high cohesiveness of a group is usually attributed to the attractiveness of the leading male to females (Yamagiwa et al. 2003a; but see Stokes 2004). Adult male–adult female relationships are considered to be the “core” of the social group; they vary depending on kinship, length of tenure, and reproductive status. Male aggression to females is common and often can be regarded as “courtship aggression.” It is higher if the female is in estrus; females usually respond submissively (Robbins 2003).

Wrangham (1979) suggests that a gorilla group represents a “permanent consortship.” Adult females usually prefer to mate with the leading male and subadult females with subordinate males (Robbins 1999). Females in multimale groups often copulate with more than one male, despite interference by dominants (Sicotte 2001); this may confuse paternity and induce all the males to protect the infants against infanticide.

If a mountain gorilla mother transfers between groups while she has a baby, if a dominant male dies, or if another silverback male takes over the group, the baby is frequently killed by the new male (Fossey 1984; Watts 1989). Infanticide causes 37% of infant deaths in the gorillas of the Virunga Volcanoes. Recently, evidence for infanticides was reported for Kahuzi-Biega too; possibly it had been suppressed earlier because the males of neighboring groups were related (Yamagiwa and Kahekwa 2004). Infanticide is suspected also in two cases in western lowland gorillas, when infants disappeared after transfers (Stokes et al. 2003).

12.5.5 Female–female relationships

Dominance relations between females are weak, their rank depending on factors such as how long they have been in the group (Watts 2001; Kapsalis 2004).
Females in large groups sometimes harass new immigrants, and aggression between females is not unusual (Watts 1996).

As females usually leave their natal group to join another group, it was long assumed that all females in gorilla groups are unrelated. In mountain gorillas, however, more than 70% of the females spend at least some of their reproductive careers with female relatives. Female maternal relatives are close associates, often interacting affinitively and supporting each other (Watts 1996, 2000, 2001).

12.5.6 Male–male relationships

As between females, there are few affiliative interactions among adult males. If the leading silverback does not tolerate another adult male in the group, he will push maturing sons to the fringe of the group, but severe contact aggression between males within a group is not usual (Sicotte 1994).

Patrilineal kinship is important for mountain gorilla males because they often remain in their natal group instead of dispersing. There they assist their fathers in resisting threats or incursions from extragroup males (Robbins 1995). Mature followers and older adolescents cooperate with dominant males against extragroup males (Watts 2000). Dominant males try to prevent these followers, however, from mating with females other than the dominants’ daughters. Coalitions of brothers may occur but are probably uncommon (Watts 1996).

12.5.7 Intergroup interactions

The home ranges of various gorilla groups and of lone silverback males overlap, so encounters are frequent, but the leaders of breeding groups generally avoid them. In mountain gorillas on average about one encounter per month is observed; it often includes aggression between silverback males. They try to drive competitors away by displaying or attacking (Yamagiwa 1987; Watts 1991; Robbins 2003). During such encounters, extragroup matings have been observed (Sicotte 2001).

Western lowland gorilla males generally show less contact aggression during encounters than mountain gorillas. They often meet in swamps or fruiting trees; the groups may feed together without hostility (Tutin 1996). They occasionally intermingle peacefully and even may nest together for one night (Kuroda et al. 1996; Doran and McNeilage 1998; Bermejo 2004). In Lossi, silverbacks often react aggressively to lone males, but usually they tolerate other groups—their reaction varies, however, and depends on the identity of the silverback (Bermejo 2004).
12.5.8 Discussion: genus Gorilla

Differences in ecological conditions seem to be more important in determining the social system of a gorilla population than its taxonomy. Mean group sizes seem to be similar across the genus Gorilla (Table 12.2), but in populations with a highly folivorous diet, the groups may become larger (Goldsmith 2003).

There seems to be a considerable influence of fruit availability on ranging and grouping patterns. During the fruiting season, large groups travel farther than small groups in western lowland gorillas as well as Bwindi gorillas. To reduce competition, the groups may spread more widely during feeding and/or form temporary foraging subgroups in western lowland gorillas and Grauer’s gorillas (Tutin 1996; Remis 1997; Parnell 2002; Goldsmith 2003; Yamagiwa et al. 2003a), which may feed and even nest more than 1 km away from the rest of the group (Bermejo 2004).

Parnell (2002) assumes that a high proportion of multimale groups in some populations may have developed because solitary males face unusually high odds against establishing a group and the current demographic conditions deter maturing males from emigration. As multimale groups provide better protection from infanticide, females may prefer groups with more than one male in the future (Watts 1996, 2000; Yamagiwa et al. 2003a; Yamagiwa and Kahekwa 2004). Dominant males may tolerate the presence of mating activities of younger males to reduce the risk of later infanticide (Robbins 1995). Nevertheless, infanticide may be very rare or absent in gorilla populations with one-male groups.

Maryanski (1987) introduced the hypothesis that gorillas, like chimpanzees, live in an “open-group system”: several gorilla groups and lone males share a home range where they meet preferred groups, socialize, and then depart. There is no evidence that a higher unit like the chimpanzee community exists in gorillas, but in some gorilla populations there do seem to be strong ties between different groups. In western lowland gorillas, males may form a dispersed male network; the males of neighboring groups are related and competition is reduced (Bradley et al. 2004). Yamagiwa and Kahekwa (2004) observed group fission in Kahuzi-Biega, which resulted in several neighboring groups with related males.

It seems that two types of association among related males evolved in gorillas: association within a group and tolerance between males of neighboring groups. The occurrence of infanticide may promote the former, and its absence may promote the latter. The social structure of gorillas may be very flexible and allow them to choose either type of social organization—even in the same population (Yamagiwa and Kahekwa 2004).

However, a network of male relatives is not a general feature of gorilla social organization. Jeffery et al. (2004) found the contrary in Lopé. Females, also
between groups, are more closely related than males between groups (males within groups are closely related). This would mean that males disperse further than females.

12.6 Chimpanzees (genus Pan)

Although the two species of the genus Pan (the “common” or robust chimpanzee and the “pygmy” or gracile chimpanzee, or bonobo) share many characteristics—physical as well as social—there are marked differences. These presumably result from their separation by the Congo River. One of the common characters distinguishing them from the other apes is a low degree of sexual dimorphism in body size, but instead a large swelling of the ano-genital region in females, and large testicles and penises in males.

12.6.1 Chimpanzee (Pan troglodytes)

12.6.1.1 Ecology

Chimpanzees have been studied at a number of sites (Table 12.1). They live in a wide variety of habitats in tropical Africa, from rain forest to closed and open forest, gallery forest, open savanna and grassland, as well as montane rain forest up to 2400 m. In Ndoki, Poulsen and Clark (2004) found the highest density in swamp forest. Chimpanzees are mainly frugivorous. Their diet consists of 48%–82% fruit, the rest consisting of leaves and other plant parts, and also more animals than other great apes, including insects as well as vertebrates, which are hunted communally (Goodall 1986; Chapman et al. 1994). The number of plant food items ranges from 55 to 328, depending on the habitat. Their diet varies seasonally, and this results in seasonal body weight fluctuations (Nishida 1990; Tutin and Fernandez 1993; Basabose 2002).

The home ranges (or territories) of chimpanzee communities vary according to habitat, season, community size, and the risk of encountering neighboring communities. The mean size is 21.6 km². In open landscapes, where food is dispersed widely, the density is very low and the home range extraordinarily large—up to 560 km². Home ranges of neighboring communities overlap (Yamagiwa 1999; Boesch and Boesch-Achermann 2000).

Within the community’s home range, each adult has his/her own core area. Most females show strong fidelity to an area once they settle there as an adult (Williams et al. 2002b, 2004). Males have larger home ranges than females.
The mean daily travel distances of individuals is about 3 km (Doran 1997; Boesch and Boesch-Achermann 2000).

In Taï, leopards attack chimpanzees; at other sites, lions prey on them (Tsukahara 1993; Boesch and Boesch-Achermann 2000). About 50% of the day chimpanzees stay above ground level. They spend the night in nests that are usually built in trees up to 50-m high (Poulsen and Clark 2004), although terrestrial nesting has also been observed in some areas, for example in Guinea (Koops et al. 2004).

12.6.1.2 Life histories and dispersal

Chimpanzee infants are nursed for about 3–4 years. The interbirth interval is usually 4–7 years (Nishida et al. 1990; Boesch and Boesch-Achermann 2000). Females may first conceive at about 9–11 years.

The mean weight of adult male *Pan troglodytes troglodytes* is 53 kg, of females 43.8 kg; *Pan troglodytes schweinfurthii* males weigh 39.5 kg, females 30 kg; *Pan troglodytes marungensis* males weigh 40.5 kg and females 32.9 kg (Groves 1986, 2001); but even within each subspecies, there is great variation both within and between populations.

The maximal age of wild chimpanzees is not yet very well known. Boesch and Boesch-Achermann (2000) assume that they may reach 50 years. In captivity, they have lived for more than 70 years.

In most populations, females usually leave their natal groups upon maturity. At Gombe, most or all adolescent females visit other communities and some may even conceive there—but only 50% of them emigrate permanently, the others returning to their natal communities (Pusey et al. 1997). In Taï, on the other hand, almost all females transfer (Boesch and Boesch-Achermann 2000). In Mahale, the transfer process lasts from 6 months to 2 years, while the females associate and mate with the males of the two communities, and 13% of the females transfer more than once there (Nishida et al. 1990). Male chimpanzees do not emigrate and cannot migrate between communities (Goodall 1986). Nevertheless, captive chimpanzees can be induced to accept new males into their group (Wilson and Wrangham 2003).

12.6.1.3 Size and structure of social units

Chimpanzees live in fission–fusion groups within their communities. They have two levels of social unit: the smaller association unit is the party or
subgroup—temporary and very variable—and the higher-level unit is the (stable) community or unit-group (\textit{Table 12.3}). Members of a community meet occasionally (fusion) and travel for longer or shorter periods in parties until they separate again (fission). On average, a party stays constant in size and composition for 24 min in Taï, in Gombe for 69, in Bossou for 126, and in Budongo for 14 min (Boesch and Boesch-Achermann 2000).

\textbf{Table 12.3}

\begin{table}
\begin{center}
\begin{tabular}{|l|c|c|c|c|}
\hline
Population & Mean party size & Range & n & Community size \\
\hline
\textit{Pan troglodytes marungensis} & & & & \\
Gombe, Tanzania & 5.6 & 498 & 50 & Boesch (1996) \\
Mahale, Tanzania & 6.1 & 218 & 29 & Boesch (1996) \\
Kibale, Uganda & 10.3 & 1–47 & 827 & 140 & Basabose (2004) \\
Budongo, Uganda & 6.3 & 1–30 & 1,824 & 46 & Newton-Fisher et al. (2000) \\
\hline
\textit{Pan troglodytes troglodytes} & & & & \\
Ndoki, Congo & 7.0 & 32 & & Malenky et al. (1994) \\
Mt. Alen, Río Muni & 4.7 & 2–7 & 3 & Jones and Sabater Pi (1971) \\
Mt. Okoro Biko, Río Muni & 11.2 & 4–23 & 5 & Jones and Sabater Pi (1971) \\
\hline
\textit{Pan troglodytes verus} & & & & \\
Taï, Côte d'Ivoire & 8.3 & 1–41 & 2,912 & 76 & Boesch (1996) \\
Taï (during fruit scarcity) & 5.75 & 395 & & Doran (1997) \\
Bossou, Guinea & 4.0 & 1–9 & 426 & 20 & Sakura (1994) \\
Assirik, Senegal & 5.3 & 267 & 28 & Boesch (1996) \\
\hline
\textit{Pan paniscus} & & & & \\
Lomako & 4.33 & 1–8 & 87 & 10 & White (1988) \\
Lomako & 7.15 & 2–17 & 26 & 22 & White (1988) \\
Lomako & 7.9 & 1–50 & 268 & >50 & Badrian and Badrian (1984) \\
Wamba & 16.9 & 1–54 & 147 & 58 & Kuroda (1979) \\
\hline
\end{tabular}
\end{center}
\end{table}

In Gombe, the average party size is 3.5 for females only, 10.7 for mixed parties, and 4.0 for males only. Single-sex parties are significantly smaller than mixed-sex parties and parties with more estrous females contain more males (Williams et al. 2002a). Estrous females are more gregarious than other classes, and they are especially associated with males (Goodall 1986; Pepper et al. 1999).
In several populations, nursery parties—several females with their infants—have been observed. In Gombe, females spend 65% of their time alone or with their offspring, in Kibale even 70%, while in Taï they are alone only 18% of their time when fruits are abundant (Wrangham et al. 1996; Pusey et al. 1997; Lehmann and Boesch 2004). During fruit scarcity, their day range is reduced and the mean party size decreased (Doran 1997). In Mahale and Kibale, food availability and the number of estrous females is positively correlated with party size (Mitani et al. 2002), but in Budongo, Newton-Fisher et al. (2000) found no positive correlation, and Basabose (2004) found in Kahuzi-Biega that fruit abundance per se does not affect party size but seasonality and fruit distribution do.

Party size is also determined by their function. During hunts for vertebrate prey, such as monkeys in Kibale, Watts and Mitani (2002) found a significant positive relationship between hunting party size and the number of kills per hunt. Success also increases with the number of males per hunting party at Gombe and Taï.

Party size depends also on community size; in large communities parties occasionally are larger than a whole small community. Therefore, Boesch and Boesch-Achermann (2000) suggest that relative mean party sizes should be compared. According to their calculation, chimpanzees have a relative mean party size of 9%–21% of the community size.

The community size may lie between 20 and 150 members. It must contain at least one adult male, but a higher number of males is usual, often more than ten. It seems that small communities retain a fission–fusion structure, but this loses much of its flexibility and the parties remain stable for much longer periods of time than in larger communities.

12.6.1.4 Male–female relationships and mating strategies

Relationships between male and female chimpanzees are usually not very close. Grooming between them, for example, is rather infrequent compared with male–male grooming. Constant and frequent proximity is particularly found in mother–son dyads.

As females copulate more often with continuously affiliative males, males interact with anestrous females to increase the chance of mating when they are in estrus (Matsumoto-Oda 2002). In Gombe, consortships have been observed in all males and 25% of conceptions occur during consortships (Constable et al. 2001). In Mahale, however, they are very rare and only 8.3% of conceptions are the result of consortship (Hasegawa and Hiraiwa-Hasegawa 1990), while in Taï
only one offspring was conceived during consortship; half of the males and 56% of the females are never seen to consort.

Prime males dominate all adult females and often try to monopolize them. Estrous females most frequently stay around the alpha male in Mahale (Takahata 1990a). In Gombe, the alpha male is responsible for 36%–45% of all conceptions and high-ranking males for 50%; in Taï, 71% of all infants are sired by high-ranking males (Constable et al. 2001).

According to Williams et al. (2002b), male aggression in boundary areas forces the females to be members of their community by settling in the center of their home range. Male coercion of females is an important element, and violence toward unfamiliar females near the edges of the defended range is particularly fierce. Nevertheless, estrous females sometimes disappear for a few days and may make temporary visits to neighboring males (Boesch and Boesch-Achermann 2000). In Gombe, 13% of copulations are with males from other communities (Goodall 1986). These extra-community matings do not very often result in conception: in Taï, extragroup paternity was found only for 7% of the offspring (one infant), and in Gombe all tested offspring were sired by males of the same community (Constable et al. 2001; Vigilant et al. 2001).

Infanticide has been observed in several chimpanzee populations, especially in Mahale (Nishida et al. 1990). In Taï and Gombe, infanticide and cannibalism by females was observed (Pusey et al. 1997; Boesch and Boesch-Achermann 2000). In Gombe and Mahale, more cases of infanticide were recorded within the community than between communities. These cases do not provide any evidence that infanticide is a successful male reproductive strategy in chimpanzees (Wilson and Wrangham 2003).

12.6.1.5 Female–female relationships

High-ranking females are the most social with other females, low-ranking females are the least. This suggests that contest competition is an important aspect of female association patterns. In Mahale and Gombe, immigrant females experience aggression from resident females (Williams et al. 2002a, b).

But females may also have affiliative relationships. In Taï, close female associations (friendships) can last for years and are very stable; some pairs spend up to 79% of their time together. According to Boesch and Boesch-Achermann (2000), higher intrasexual competition and higher involvement in the social interactions of males make it profitable for females in Taï to develop long-term friendships with other females and to form stable alliances. Maternal kinship is assumed to play a strong role in female affiliation (Kapsalis 2004).
12.6.1.6 Male–male relationships

Male chimpanzees associate more strongly with one another than do females with other females and males with females. They form coalitions in all populations studied (Boesch and Boesch-Achermann 2000; Newton-Fisher 2002), and apart from coalitions, friendship between males has also been observed (Nishida and Hosaka 1996). Which individuals form affiliative relationships is not clear; genetic studies in Kibale showed that maternal kinship is not strongly associated with male–male association (Kapsalis 2004).

Among the males, there is a linear dominance hierarchy, and rank reversal generally results from dyadic fights (Takahata 1990b; Muller 2002). The alpha male is the most active groomer; he tends to move first and be followed by subordinates (Takahata 1990b). In Tāi, the leader of a community announces his presence by drumming; this also gives information to other individuals about the direction and speed of group movement (Boesch and Boesch-Achermann 2000).

Agonistic confrontations between males are observed regularly, and they are most aggressive between the two highest-ranking males. Coalitions in attacks are frequent; in Tāi it is mostly low-ranking males coalescing against dominant individuals (Boesch and Boesch-Achermann 2000). Coalitions are also formed for hunting, for the intragroup control of widely dispersed females and to monitor territorial borders (Stanford 1998).

12.6.1.7 Intergroup interactions

Most interactions of chimpanzee males with neighboring communities involve only auditory contact—pant-hoots, a long distance call. These pant-hoots are also used to advertise their presence and numerical strength. Males almost always show fear or hostility to strange males (Wilson and Wrangham 2003).

Chimpanzee males invest considerable time and energy in defending the home range of their community or locating their neighbors; the home range is controlled by groups of at least four males on a weekly basis in Tāi (Boesch and Boesch-Achermann 2000). During those patrols, they remain silent and actively search for signs of the neighbors. They make incursions into the home ranges of the neighbor communities, sometimes of more than 1 km, and if they encounter strange males, they attack them. Not only males are attacked but females too, except for tumescent females (Pusey 2001; Williams et al. 2004). Females with or without infants often join attacks, but they tend to avoid direct physical contact with the other community.
Hostile intercommunity relations have been observed at all sites. Intraspecific violence is one of the leading causes of mortality for eastern chimpanzees (Wilson et al. 2004). At Gombe and Mahale, the destruction of a small community by a larger one, including systematic attacks and killing of individuals by males from a larger community, has been observed. Wilson and Wrangham (2003) provide a good overview of such intercommunity conflicts. So far, there is no consistent evidence from the field, that the communities find more or better sexual partners and new resources as a result of the fights (Boesch and Boesch-Achermann 2000). Extensive female transfer after a violent fight between communities was observed only in Mahale (Wilson and Wrangham 2003). In Gombe, adult parous females join other communities only when all males of their community have been killed.

Williams et al. (2004) conclude that male chimpanzees cooperatively defend territories that contain food resources for themselves, their long-term female mates, and their offspring, and they try to extend the size of the community’s home range because a larger area means greater availability of food and higher female reproduction. Infanticide during intercommunity encounters can also be interpreted as the removal of future competitors. Concerning intercommunity killings by adult males, data from various study sites most strongly support the hypothesis that attackers reduce the future coalition strength of rival communities.

12.6.2 Bonobo (*Pan paniscus*)

The most important sites where bonobos have been studied are listed in Table 12.1.

12.6.2.1 Ecology

The typical habitat for bonobos is the lowland rain forests and swamp forests of the Congo Basin. In some areas, they also live in dry forest and visit grassland. They eat up to 147 food items; 72%–90% of their diet consists of fruits (Kano and Mulavwa 1984; White 1992; Yamagiwa 2004). The amount of meat consumption is not as high as in some chimpanzee populations but seems to fall within the general range of chimpanzees. Bonobos (including females) hunt small mammals, usually solitarily (Fruth and Hohmann 2002).

In Wamba as well as in Lomako, the home ranges of communities overlap extensively (Idani 1990; Hohmann and Fruth 2002). Their size lies between
22 and 58 km² in Wamba (Idani 1990). Each adult has an individual home range or core area within the community’s home range (White 1996).

Bonobos may experience lower leopard predation pressure than chimpanzees because they spend more time off the ground (Boesch 1991). Outside the forest, they seem to be very careful; if they feed on fruit in the grassland, they remain quiet (Myers Thompson 2002).

12.6.2.2 Life histories and dispersal

Bonobo infants are weaned at 3–4 years of age, and the interbirth interval is about 4–7 years (Lee 1999; Yamagiwa 2004). Females conceive for the first time at about 10–14 years. Adult males have a mean weight of 39.2 kg, females weigh 31.5 kg (Groves 1986).

Females transfer to other communities as older juveniles or early adolescents (Furuichi 1989). Paternity analyses suggest that there must be a large exchange of females between communities (Gerloff et al. 1999). Males tend to stay in their natal community. Occasionally, they may transfer to other communities, but this is rare (Hohmann 2001).

12.6.2.3 Size and structure of social units

Much like common chimpanzees, bonobos live in a fission–fusion social system. Parties usually contain mature individuals of both sexes with more females than males. The proportion of all-female parties in Lomako is high, of all-male parties low (Hohmann and Fruth 2002). If estrous females are present, the proportion of males increases (Hohmann and Fruth 2002). Lone individuals are rare—usually males travel alone (White 1996).

Party sizes are determined by food availability: if more fruits are available and if the food patch is large, the parties grow larger. Males disperse when food becomes scarce but females do not (White 1998). As bonobo food includes herbaceous plants that are abundant in the rain forest during all seasons, feeding competition is low.

In general, bonobo parties are large in Lomako and Wamba, compared to chimpanzees. While chimpanzee parties are 9%–21% of the community size, bonobo parties consist of 21%–89% of the community. Bonobo parties last longer than those of the chimpanzees at Taï and Gombe (in Wamba 86 min, in Lomako 102 min; Boesch and Boesch-Achermann 2000).

Community sizes in Wamba are very variable, ranging from 33 to more than 100 members (Idani 1990). The cohesion of community members is high, and
they stay together most of the time. In Lomako, several parties may congregate in the evening to nest in proximity to each other (Hohmann and Fruth 2002). Community members may be separated by kilometers for days or weeks (White 1996).

12.6.2.4 Male–female relationships and mating strategies

In bonobo communities, either females are dominant over males or both sexes are codominant/egalitarian (Gerloff et al. 1999). Long-term bonds are found predominantly between heterosexual dyads and involve not only close kin but also unrelated individuals. Relatives associate and groom more often, however, and kinship ties are important between males and females. The highest association rates are observed between adult females and their adult sons: males receive agonistic aid from their mothers in conflicts with other males (Hohmann et al. 1999; Kapsalis 2004). Aggression by males toward females is less intense than in chimpanzees. Females may form alliances to attack males (Furuichi 1989; Hohmann and Fruth 2002).

Bonobo mating is opportunistic and promiscuous and involves no or little aggression between males. The maximal swelling lasts for a large proportion of the cycle, therefore males establish long-term bonds with females that exceed tumescence (Fruth et al. 1999). Nevertheless, high-ranking males have a strong tendency to monopolize tumescent females and they sire more offspring (Kano 1996; Gerloff et al. 1999).

Extra-community copulations are not uncommon, and females are rarely prevented from mating with members of neighboring communities. The number of infants sired by nonresident males is low; more than 80% of the infants in Lomako are fathered by resident males. No infanticide was observed so far in bonobos (Fruth et al. 1999; Gerloff et al. 1999).

12.6.2.5 Female–female relationships

Female bonobos are more affiliative and cohesive with each other than chimpanzees. Contact frequencies between females are higher than between females and males or between males. They associate and forage in larger parties for most of the year, share food and support each other in food defense (Hohmann and Fruth 2002). These affiliative bonds are not particularly observed between related females; female associations are not based on kinship (Kapsalis 2004).

Female bonobos groom less than male–male and male–female dyads but show a unique behavior called genito-genital rubbing, especially in the context of feeding: two females embrace each other ventro-ventrally and rub their genital
swellings together with rapid sideways movements. The function of this behavior was discussed by various authors, such as Hohmann and Fruth (2000), who observed genito-genital rubbing six times as often as female–female aggression. According to their analysis it serves reconciliation and tension regulation.

12.6.2.6 Male–male relationships

Although strong bonds between males exist, especially in Wamba, they are less prominent than the bonds among females (Hohmann and Fruth 2002). Unlike chimpanzee males, bonobo males have even fewer contacts with other males than with females (White 1998). High-association rates are observed between maternally related adult brothers (Kapsalis 2004). Alliances are unusual between males (Hohmann et al. 1999). The males establish dominance relationships with each other, but aggression is less intense than in chimpanzees and conflicts are often settled in a nonagonistic way (Hohmann and Fruth 2002).

12.6.2.7 Intergroup interactions

Bonobo communities do not seem to search for and contact neighboring communities. Lomako males have never been seen to make border patrols (Hohmann and Fruth 2002). In Wamba, intergroup encounters vary from group fights to peaceful intermingling. In general, encounters are peaceful and communities may spend hours together. Females take the initiative in the temporary fusion of communities. During these community meetings, males keep a certain distance from the males of the other group. The most prominent form of intergroup interaction between males and females is copulation, and relations between resident and unknown females are characterized by friendly contacts (Idani 1990; Kano 1996; Gerloff et al. 1999).

There are frequently aggressive interactions between males when they approach, but direct body contact and cooperative attacks are rare; the aggressive interactions are never as fierce as those reported for chimpanzees (Idani 1990; Hohmann and Fruth 2002). Agonistic aid during conflicts between members of different communities has never been reported. Severe aggression does occur, however, when mixed-sex parties encounter unknown males; in such a case, the strangers are charged by the males and also by the females (Hohmann et al. 1999). No fatal aggression was so far observed between bonobo communities in Wamba (Kano 1996).
12.6.2.8 Discussion: genus *Pan*

Usually the chimpanzee social system has been regarded as male-bonded, with strong kinship ties between the males of a community but no relationships between the females. Experience from various field sites does not always support this idea and indicates that it is much more complicated and variable. In Taï, males within a community are on average not significantly more related than females, and the group members have more relatives within their home community than outside (Vigilant et al. 2001). Association patterns do not support the view of strong bonds between males in general (Pepper et al. 1999).

Taï chimpanzees may be bisexually bonded, while other populations are male-bonded, and more cooperation is found in Taï than in eastern chimpanzees. The reasons are presumably differences in habitat. Boesch and Boesch-Achermann (2000) assume that the forest environment allows or forces bonobos and chimpanzees to build larger and more cohesive parties. Doran et al. (2002b) hypothesize that permanent female association with males is a female counterstrategy to infanticide risk and that more infanticide occurs in habitats with considerable annual variance in fruit production. Bonobos live in a still more stable environment than Taï chimpanzees—this may lead to even more stable party sizes (Doran 1997). Bonobo parties seem to be large compared to chimpanzees (∙ Table 12.3), but Hohmann and Fruth (2002) state that Lomako bonobos are within the range of variation of chimpanzees.

Bonobo communities seem to be composed of unrelated females who are highly affiliative with each other and related males who are not highly affiliative with each other; with homosexual behavior females directly control competition (Boesch and Boesch-Achermann 2000). Aggression between males and between the sexes is less intense than in chimpanzees and conflicts are often settled in a nonagonistic way. Bonobos in general have more relaxed relationships than chimpanzees that do not depend on kinship, as paternity studies show that there is no matrilineal organization (Gerloff et al. 1999).

The typical chimpanzee/bonobo social structure is a multimale group with a fission–fusion structure. Similarities are obvious with respect to party size and association patterns. Female bonding in bonobos does not exceed that of some chimpanzee populations; differences between the two species are the proportion of female party members and the frequency of mixed parties (Hohmann and Fruth 2002). Chimpanzees as well as bonobos have the potential for great social variability, with considerable capacity for cooperation, reciprocal interactions, and coalitional behavior (Boesch and Boesch-Achermann 2000).
Despite the common basis, the two species show some differences in social behavior. Wrangham et al. (1996) think that this can partly be explained by the differences in feeding competition: chimpanzees and gorillas live sympatrically in many areas while bonobos do not have a great ape competitor.

### 12.7 Sympatric ape populations

Sympatric apes share a great part of their diet—in Asia as well as in Africa. This is especially visible in fruits. Sugardjito et al. (1987) observed some competition between orangutans and siamangs in Gunung Leuser, and one benefit of grouping for Sumatran orangutans may be that siamangs cannot drive the youngsters away from fruiting trees. More obvious, however, is the interspecific competition between chimpanzees and gorillas in Africa.

The dietary overlap between gorillas and chimpanzees ranges from about 50% at Kahuzi-Biega to 60%–80% at Lopé and Ndoki. In Kahuzi-Biega, all fruit species eaten by gorillas are also eaten by chimpanzees. Overt interspecific competition between chimpanzees and gorillas has not been observed at any site; instead, competition avoidance is commonly seen (Kuroda et al. 1996). Interspecies relationships are more peaceful than intergroup relationships within the two species (Yamagiwa et al. 1996, 2003b). In Gabon, Okayasu (2004) observed close interactions between gorillas and chimpanzees; occasionally the groups would mix and play and even sleep at the same site.

During fruit scarcity, gorillas increase the proportion of herbaceous vegetation in their diet, while chimpanzees as obligatory frugivores continue to search for fruit. The two species obviously found different niches (Yamagiwa et al. 1996, 2003b), and some habitats are used almost exclusively by one species (Tutin and Fernandez 1993; Malenky et al. 1994; Kuroda et al. 1996; Rogers et al. 2004).

Kuroda et al. (1996) suggest that the low population densities of gorillas and chimpanzees in Lopé and Kahuzi-Biega might partly be due to competition. Possibly interspecific competition over food affects foraging strategies and may have caused divergence in grouping patterns. The larger party sizes of bonobos are possible because of the high density of terrestrial herbaceous vegetation; as gorillas mainly eat these plants, sympatric chimpanzees may be forced to take a different foraging strategy and to form smaller parties (Wrangham et al. 1996; Yamagiwa and Takenoshita 2004). The effects of competition have not been analyzed yet, but they are difficult to study—also because additional competitors like elephants have to be considered (Rogers et al. 2004).
12.8  Conclusions and the genus *Homo*

Although some great ape populations have been studied for decades, their social systems are not yet completely understood. The Asian apes seem to be less social than the African apes; this may be due to food types and distributions in Southeast Asian forest, which may differ strongly from African forests. All great apes lead “individual-centered lives,” but they need the interaction with familiar conspecifics. Despite their tendency to congregate, their social structure is characterized by weak ties, compared to female philopatric primates.

Female transfer is common to all species. They have a tendency toward fission–fusion grouping; females lack sharply defined dominance relations, and intrasexual bonds among nonkin can be relatively strong. It is difficult to assign a social system to each ape species (or to the family Hominidae in general) because there is remarkable intraspecific variability in social organization and structure. Especially frugivory requires a mobile and flexible population.

Compared to the great apes, humans show an even greater variability in social structure—nevertheless, there are certain trends across all human societies (Rodseth et al. 1991): males maintain consanguineal kin ties; females maintain consanguineal kin ties; males cooperate in conflicts against other males; females also cooperate but rarely in physical conflicts with other females.

According to Knauft (1991), simple human societies are decentralized, and there tends to be active and assiduous devaluation of adult male status differentiation. Among complex hunter-gatherers and with the advent of sedentism and horticulture/agriculture, male status differentiation increased. There seems to be a similarity between great apes and middle-range human societies in terms of competitive male dominance hierarchies. Such dominance relations may not be particularly adaptive in environments of low resource density and predictability; this may have led to the simple egalitarian hunter-gatherer societies that nowadays live in extreme environments.

Most human societies are characterized by female-biased dispersal and male philopatry. Long-term pair bonds between males and females are common, although their form, strength, and duration vary between societies. Moreover, these bonds are not identical with mating and grouping patterns (Pusey 2001). There has been much speculation on the “natural” human mating system. Although fossils of our early ancestors show extreme sexual dimorphism, modern human males are only about 15% larger than females; the relative size of testes in humans is much smaller than in chimpanzees and comparable to “monogamous” or one-male group species. This means, theoretically, that we are a frequently
mating species with low variance in male mating success (Wrangham 2001). Polygamy with only some males producing many offspring thus cannot be the common mating system in humans; but social monogamy is not common either (Low 2003).

The social system of humans certainly has several levels—like the social system of Pan. Dunbar (1993) developed the hypothesis that there is a species-specific upper limit to group size that is set by cognitive constraints. This would mean that human groups can be much larger than those of the great apes. According to Dunbar, group size depends on the maximum number of individuals with whom an individual can maintain personal contact. He discerns (in modern hunter-gatherer societies) the group levels overnight camp (30–50 members), band/village (100–200 members), and tribe (1,000–2,000 members). Dunbar’s overnight camp certainly is not the smallest human grouping above the individual. Rodseth et al. (1991) and Pusey (2001) state that the majority of human societies consists of conjugal families united in stable communities—but also relatively autonomous families.

But what is the central, stable component of the human social system? Even ape specialists have contradictory opinions. De Waal (2001) thinks that the nuclear family is the basic social grouping of humans and that this unit is unique to our species, although Low (2003) states that it is rather unusual in human societies. Perhaps the nuclear family is an especially successful social structure in modern industrialized societies. Ghiglieri (1989) calls the social structure of humans a multimale kin-group, a stable, semiclosed fission–fusion community.

Acknowledgments

I am very grateful to Colin Groves who improved the manuscript in many respects.

References


Bermejo M (2004) Home-range use and intergroup encounters in western gorillas (Gorilla...
g. gorilla) at Lossi Forest, North Congo. Am J Primatol 64: 223–232


Jones C, Sabater Pi J (1971) Comparative ecology of Gorilla gorilla (Savage and Wyman) and Pan troglodytes (Blumenbach) in Rio Muni, West Africa. Karger, Basel


Kano T, Mulavwa M (1984) Feeding ecology of the pygmy chimpanzees (Pan paniscus) of
Kummer H (1971) Primate societies. Aldine Atherton, Chicago


Schaller GB (1963) The mountain gorilla. Chicago University Press, Chicago


Williams JM, Pusey AE, Carlis JV, Farm BP, Goodall J (2002b) Female competition and male territorial behaviour influence female chimpanzees’ ranging patterns. Anim Behav 63: 347–360


Brain size has traditionally been employed as a measurable proxy for species intelligence. Using allometric scaling of brain size relative to body size shows the biological cost suffered from investment in brain tissue. Shifts in diet type are the engine permitting increased investment in brain tissue because higher energy diets allow a larger brain at any given body size. Relative brain size, however, confounds effects of gut size required for particular diets with effects of brain size required for enhanced cognitive function. In contrast, the absolute size of brain parts specialized for particular functions gives evidence of the computational power of those systems. Correlational analyses strongly imply that demands of social complexity, rather than difficulties associated with frugivory or embedded foods, led to evolutionary increase in simian primate brain size. Primate brain expansion has largely involved neocortex, and among living primates, neocortex size predicts frequency of use of tactical deception and of innovative responses. These capacities likely rely on extensive memory for social information, but there is evidence (only) among great apes for understanding how systems work, whether social or technical. Representational understanding may derive from the ability to parse complex behavior, allowing imitative learning of elaborate new skills.

13.1 Introduction

There are three ways in which physical anthropologists might be interested in primate intelligence, corresponding to different theoretical formulations. Firstly, a very widespread view among social scientists is that human intelligence is unique and indivisible, shared with no other species (Macphail 1982, 1985). This does not mean “unique as each species’ intelligence is unique,” nor is the claim simply that it is “uniquely large”; rather, human intelligence is incomparable, and anyone seeking seriously to relate animal and human intelligence is misled. As regards origins, human intelligence would be taken to be a consequence of developing the faculties of language and speech, traits also unique to humans. Thus, on this stance, the key question for human evolution is at what point in human evolution did our makeup change in a way that permitted the development of language:
the (unique) origin of intelligence equates to the (unique) origin of language. Replacing a hard question with a famously intractable one leads to no obvious avenue for further research. Moreover, while in the end it may turn out that the gulf between human intelligence and that of any other animal is an unusually large one, to assume a priori that there is no comparison makes good sense only for a creationist. I will not therefore consider this position further.

At the opposite extreme, intelligence has been treated as a continuously varying quantity: traditionally called “g” in psychometric psychology, for “general intelligence,” or IQ in common parlance. On this view, the evolutionary issue is whether intelligence changed gradually or in steps, and we can use comparative phylogenetic data of primates to pin down when in human evolution such changes occurred (see Chapters 4, 23 in Volume 1). Further, by exploring the ecological correlates of intelligence in living primate species, we may gain insight into the current adaptive value (usefulness) of intelligence and thereby reveal clues to its evolutionary origin. What primatology needs to bring to this enterprise is the equivalent of an intelligence test for animals: a fair way of comparing the intellects of living species. This has been the most popular and to date the most productive approach to primate intelligence, but in this chapter I will suggest that it is only a rough approximation to the truth and may be of limited future value. Finally, lying between the two extremes of intelligence-as-entity and intelligence-as-quantity, intelligence can be treated as a heterogeneous skill package: a mixed bag of devices and processes, endowments and aptitudes, in fact all those capacities that lead us to label everyday behavior as “intelligent.” I will argue that this rather messy vision in fact offers the most promise for the future use of comparative data in understanding the evolution of our most vaunted feature, human intelligence.

13.2 Intelligence as quantity

Intelligence as measured within psychology is a matter of differences in ability among people. Strictly speaking, then, intelligence is an individual-level phenomenon. However, over and above the differences in problem-solving ability among humans (and presumably also in other species), individuals share a “common denominator” potential for intelligence by virtue of species membership. It is this genetically based, common denominator that we mean by talking of “differences in intelligence between species.” Species differences in intelligence are clearly liable to be much greater than the small differences within human intelligence.

How can species intelligence be measured? An obvious starting point is to co-opt and adapt the tools of psychometric psychology: surely measures designed
for the small variations in human intelligence will prove sensitive enough for
the larger species differences? In practice, this is not as easy as it might seem.
One reason is that, although most people consider they can deduce intelligence
from everyday behavior, psychologists have preferred to define intelligence oper-
ationally, as “what IQ tests measure.” As a result, we can draw upon sadly little
expertise in psychometrics on how to infer intelligence from behavior. Still, a
considerable battery of tests, broadly comparable to the items of an IQ test, have
been applied to animals, using the tools of the animal learning laboratory:
Skinner boxes, Wisconsin General Test Apparatus, and so on. Unfortunately,
the results have been disappointing, as far as measuring species intelligence
goes—and perhaps this should not surprise us. Human intelligence testing has
shown the inappropriateness of grabbing a test, devised in one culture, and using
it on people from another culture. The problem is that tests must be calibrated
against some other achievement (for human IQ tests, usually educational suc-
cess), and these measures of achievement are not culture free; the difficulty is
obviously far greater across species. Only a few animals are motivated by similar
rewards to ourselves, see the world in rather human ways and interact with the
world in a similar way to humans. If we rely on human estimation of difficulty in
some “behavioral IQ test for animals,” we are liable to equate cleverness with
similarity to ourselves. The history of laboratory-based comparative psychology
has gone through a series of cycles, with striking species differences in intelligence
first claimed, and then the difference later discovered to lie rather in perceptual
capacity, motivation, or species-typical traits better understood as special-
purpose adaptations for particular environmental features (see reviews by Warren
1973; Macphail 1985). Goldfish were once considered less intelligent than rats,
because they could not learn visual discrimination tasks—until it was realized
that their visual acuity was greater in a downward direction. When tasks were
presented on the bottom of the tank, the difference vanished. Among primates,
colobines were (and still often are) considered unintelligent compared to cerco-
pithecines—but the latter are frugivoros, readily motivated to perform tasks in
return for fruit or other concentrated food. Colobines are folivores, not adapted
to compete for (or even able to digest) small items of high-quality food; it is
hardly a level playing field. Barn swallows show phenomenal abilities to find their
way over great distances to return to the same barn after many months, using the
positions of sun and stars, polarized light, and magnetic fields; but they show no
other signs of great intelligence, and migratory abilities are not helpfully seen as
evidence of intellectual level. Small wonder then, that some biologists have
doubted whether intelligence is an appropriate measure by which to compare
animals at all. Animal adaptations are fascinating, they argue, but calling that
intelligence adds nothing to our understanding. Often, psychologists’ definitions
of intelligence stress the need to deal with environmental challenges: e.g., “the aggregate or global capacity of the individual to act purposefully, think rationally, and to deal effectively with his environment” (Wechsler 1944), or “the faculty of adapting oneself to circumstances” (Binet and Simon 1915). If intelligent for an animal species means “well adapted to the environment,” then presumably all species are “intelligent” in their own, nonhuman ways. And while many performances of animals look intelligent in the human sense, there is every reason to suppose that their development is under tight genetic guidance.

Surely, however, adaptations specific to solving particular environmental problems should be distinguishable from real intelligence by its quality of flexibility, allowing individuals to find their own solutions even to novel problems? Species-typical performances are most likely adaptations (although not all will prove to be: the wearing of bodily adornment or coverings is, after all, species-typical in humans!). But the importance of species-level intelligence is its potential for allowing individual flexibility in learning and problem solving. This approach points to the use of observational data of natural behavior to deduce intelligence, just as we do in everyday among ourselves. However, in examining natural behavior for signs of the individual-level flexibility and creativity that can signal species intelligence, the same danger of confusing genetical adaptations with flexible intelligence occurs as with laboratory testing. An anecdote of my own error may serve as a cautionary tale. Watching border collies herd sheep, I did not doubt the dog’s greater intelligence. The sheepdog responds to the whistles of a shepherd with flawless out-maneuvering and controlling of a hundred sheep. Of course, this wonderful performance depends on the innate antipredator reactions of sheep. Bunching and running in tight-packed flocks when attacked makes it difficult for a wolf to single out a potential kill, but easy for a sheepdog to maneuver a group. The dog is also equipped with innate tactics, partly as a result of its wolf ancestry and partly as a result of thousands of years of domestication; these tactics can be seen in any untrained sheepdog let out to chase some sheep. But the dog is also able to learn the complex system of whistle signals and is then able to deploy its tactics to order; the sheep, in contrast, are unable to overcome their innate restrictions. My faith in this simple picture was shattered by spending some time with Gujarati shepherds on the Little Rann of Kutch in India. Like British shepherds, they whistle their commands; it seemed a familiar scene. Eventually, I noticed the sheepdogs: asleep. The sheep in Gujarat learn to understand the shepherds’ commands and follow them, treating the shepherd as herd leader. The dogs’ role is not one of herder but only a source of protection from wolves. Dogs may seem especially intelligent to people because they happen to use facial musculature for visual communication, giving rise to expressions that resemble our own and have similar meanings, and because their
forward-facing eyes and long nose make their direction of attention obvious. We can “see” what they are thinking, and recent work shows that domestication has equipped dogs with particular traits that fit into this two-way cooperation (Miklosi et al. 2004). Sheep are foreign to us, because they rely more on olfaction and their facial expressions are relatively cryptic to us, but it is premature to assume them unintelligent.

A proper comparison of intelligence, shown by each of the species relevant to reconstructing human cognitive evolution, is therefore a tall order. Attempts to use general-purpose laboratory tests have foundered on extraneous differences in natural aptitudes, perceptual capacities, and motivation; in any case, relatively few species are available for detailed examination in captive settings, and comparative phylogenetic analysis depends on using a broad range of species. Yet to accurately attribute differences in natural behavior to intelligence, rather than other evolved aspects of the species biology, requires in-depth study of each species under a range of conditions and so is almost as restricted in what data are available. In consequence, most progress in evaluating animal intelligence as a quantity has been made by using an indirect indication of intelligence, brain size, which can be accurately measured anatomically.

13.3 Brain size as a measure of intelligence

The clearest evidence that brain enlargement confers adaptive advantage comes from examining the costs of a large brain: species can tolerate retention of neutral traits, but for a costly organ to evolve necessitates compensating advantages. Brain tissue is metabolically expensive (Aiello and Wheeler 1995). In adulthood, the human brain consumes about 20% of the basal metabolic rate, and during childhood this percentage rises to 50%. Moreover, this demand for energy is remorseless: unlike other organs, the energy supply to the brain has to be constant, and irreparable damage results from only a few minutes of interruption. Having a large brain has incurred other disadvantages for us, as well as this energetic drain. At birth, the human child’s head is a tight fit in the birth canal compared with the easy passage of other great ape babies (Leutenegger 1982). Birth is consequently a prolonged, often painful, and sometimes dangerous process for mothers; for other great apes, birth takes only a few minutes. Finally, human brains grow for an unusual amount of time, considering their size at birth (Harvey and Clutton-Brock 1985). During this phase of postnatal brain growth, human babies are relatively immature and helpless, so require years of time-consuming care from the mother or family. Among primates, the only plausible explanation for brain enlargement in the face of such clear costs is an intellectual benefit. To measure
species intelligence, then, it should be possible to use brain enlargement. Diffi-
culties arise, however, in deciding the baseline for measuring enlargement.

Larger animals, in general, have larger brains. As the absolute size of living
things changes, the relative proportions of their parts are generally found to
change. In this case, absolutely larger animals have relatively smaller brains than
expected from linearly scaling-up smaller ones (Jerison 1963, 1973). These regular
trends have led to the use of allometry to calibrate brain enlargement, against a
baseline of the size expected from body weight. In allometric scaling, for a given

Comparing species in intellectual ability, by using their relative brain sizes
calibrated with allometric scaling, however, leads to a paradox which brings the
whole approach into question. Scaling brain size against body size is implicitly
making a strong claim about the functioning of neural tissue. The implication is
that an animal with an expected brain size of 2.0 g and a real brain weighing 2.1 g,
and an animal with an expected brain size of 200 g and a real brain weighing 210 g,
are “really” equally brainy—even though the latter differs from the expected
brain weight by 100 times as much neural tissue as the former. This is a very
puzzling result for anyone used to computational (Turing) machines, since these
are ultimately limited in power by the number of their elements. The paradox
comes from mixing metaphors of what the brain is doing. If the brain is a sort of
“on-board computer” (Dawkins 1976) that governs intelligent function, then the
absolute number of neurons available for computation must be relevant, not the
number relative to body size. (The logic here is that neural transmission speed
is known to be the same in all mammalian brains, and evolution will have
optimized neural programming in each species: thus differences will not reflect
relative efficiency or hardware or software, as is the case in most artificial com-
puters.) Bigger brains will be better brains, when it comes to flexible and intelli-
gent responses, regardless of the species’ body size. In contrast, using allometric
scaling against body size presupposes a more traditional scheme in which animal
brains function by making responses to stimuli in a more-or-less reflex manner. The underlying model of the mind is then closer to an automatic telephone exchange than to a computer. Lines from/to subscribers in the telephone system model correspond, in bodies, to sensory and motor neurons. So, input/output connections will determine how big the system to handle them needs minimally to be. Larger bodies need larger brains for these prosaic purposes, and only measuring brain tissue relative to body size will show the extent to which processing can be more flexible and intelligent than the minimum. The on-board computer and telephone exchange models cannot both be right—or rather, they will be appropriate for different systems within a single brain, and the error is to assume that one or other can be neglected entirely (Byrne 1995b, 1996b). Accepting this more complex view, those brain parts involved in non-computational body-function should increase in size in some regular way with body size, whereas those parts used for computation should not. The absolute amount of brain tissue free for computation should tell us the potential intelligence the brain can show, not the amount relative to body size. It should be no surprise, then, that allometric analysis of brain size relative to body size and alternative methods that in some way measure absolute sizes produce conflicting results (see Deaner et al. 2000).

### 13.4 Brain size relative to body size

Allometric scaling shows that, among mammals, the primate order as a whole is larger brained than most other groups (Jerison 1973). But when strepsirhine primates are partitioned from the rest, they turn out to have brains about the size predicted from mammalian body size (Passingham and Ettlinger 1974). The monkeys and apes, however, have brains twice as large as those of average mammals of their size. What does this mean? Following the arguments of this chapter, disproportionate brain size in relation to body size is not simply a matter of greater intelligence and must be understood in terms of costs as well as benefits. Any species with brain relatively large for its body size inevitably incurs greater risks than a small-brained relative, from the remorseless demand for higher metabolic energy. The relatively large-brained monkeys and apes are thus bearing a much greater cost from their larger brains, on average, than most mammals: how, and why, are these increased costs acceptable?

Primates with home ranges that are large in area tend to have relatively large brains (Clutton-Brock and Harvey 1980), and this has been used to argue that environmental complexity has a powerful influence on primate brain size. But there is an alternative explanation for that correlation: as an artifact of selection
for bigger bodies in more folivorous species (Byrne 1996b). Folivory relies on a complex or at least large stomach: for example, the foregut fermentation chamber of colobine monkeys or the large hind-gut of gorillas. Leaves are relatively abundant in most primate habitats, so primates with more folivorous diets can find sufficient food for their nutrition in smaller home ranges. By contrast, frugivory requires a larger range area, for year-round access to a variety of fruit species and other sources of nutrients, but the high sugar content allows digestion by a shorter gut. Other things being equal, primates that eat more fruit will have smaller bodies and larger home ranges than those that eat more leaves, causing a correlation between frugivory and relative brain size. Variations in brain size relative to body size may therefore be a side effect of differences in body size due to diet type rather than a direct result of selection for larger brain size.

Since gut tissue is metabolically as costly as brain tissue, and diets requiring only small guts (frugivory, meat-eating) often provide a surplus of energy by the time a nutritional balance is obtained, primates with small guts are on the whole likely to be better able to “afford” larger brains (Aiello and Wheeler 1995). Clearly, the grade shift toward meat-based diets in the later hominins might be related to the massive brain expansion in these species. The relationship between energy supply and the size of brain that can be afforded for a given body size does not, however, explain why larger brains should have evolved in some taxa and not others: the opposite is just as feasible, a trend toward smaller brains in species with less need for high-quality diets (see Figure 13.1). Moreover, the fact that the energy demands of the brain are constant suggests that the large brain/high-energy diet is a particularly risky evolutionary strategy for which a compelling competitive advantage must exist.

One possibility is that a niche for a high-quality diet specialist may happen to become available, and the survival cost of large brain size is thereby lessened for any species that exploits that opportunity, resulting in a new equilibrium at a higher relative brain size. However, higher quality diets are usually based on sparse, hard-to-find, hard-to-access, or hard-to-process foods (e.g., fruit, nuts, meat), so this is not entirely plausible. More often, the causal chain is likely to run in the opposite direction: from a real need for higher intelligence, precisely in order to exploit a new, high-quality food supply. It might even sometimes be the case that ecological pressures for greater intelligence—and thus a larger brain—drive selection for a larger body, in order to support it. The frequently noted trend over geologic time for species of the same taxa to increase in size might sometimes be a consequence of an arms race of intellectual competition, selecting for larger brains and consequently larger bodies to support them at the same level of survival risk, coevolving with changes in diet.
13.5 Absolute brain size

To measure the computational power of a brain, it will be necessary to estimate the absolute volume of neural tissue available for flexible, problem-solving purposes: the lack of any overwhelming correlation between overall brain size and observed smartness in animals suggests that simply using the total brain volume will not do. Jerison (1973) was well aware of the need and developed an index of “extra neurons” by calculating the absolute number of neurons beyond what was minimally necessary for bodily function. With a similar aim, Bauchot and Stephan (1966) attempted to identify the taxonomic group with least intelligence as a baseline, from which all others deviated. However, difficulties in deciding what volume of neural tissue is minimally necessary, or whether any species entirely lacks flexible intelligence, have prevented general adoption of these approaches. In recent years, the same problem has generally been tackled in a different way by comparing the volume of one part of the brain, the neocortex, against the rest. There are two assumptions involved here. First, that the primate brain has undergone mosaic evolution, with some parts growing in size and power at the expense of others, for a given overall size (Barton 1998; Barton and Harvey 2000). On a broad scale, this assumption is hard to doubt: for instance, haplorhine primate brains are clearly more dominated by visual cortex and less by olfactory lobes than those of most other mammals, including strepsirhines.
The second assumption is that we can safely attribute a main function to expansions of particular regions—in particular, that the neocortex is involved in abstraction, thinking, and executive functions such as effective problem solving. This is supported by a long history of deducing brain function from task failures after accidentally or deliberately inflicted lesions, somewhat problematic to interpret but consistent in pattern, and more recently supplemented by the local pattern of energy use revealed by brain imaging. (But note that rodent work indicates considerable involvement of the midbrain in executive functions.) More pragmatically, it can be noted that the increase in brain size in the primates over that of other mammal groups is chiefly due to neocortical enlargement; and among primates, it is the neocortex that varies most strikingly between species, whereas the rest of the brain shows much less evolutionary change (see Passingham 1981). This implies a strong selection pressure for neocortical enlargement in primates, and an intellectual function is the only serious candidate for this selection pressure.

If it is taken, then, that primate neocortical enlargement measures specialization in some sort of intelligence, can ecological correlates indicate the function it subserves? Sawaguchi and Kudo (1990) found that the neocortex was larger in species living in bigger social groups, both in strepsirhines and in frugivorous platyrhines. Also, in frugivorous haplorhines, polygynous species (one male living with more than one female) had larger neocortices than monogamous species. These findings hint at a social origin of intelligence (Jolly 1966; Humphrey 1976; Byrne and Whiten 1988), and Dunbar (1992, 1995, 1998) has gone further to support this idea, examining whether neocortex size correlates with measures of social or environmental complexity in living primates. He used both raw neocortical volume and several more complex functions; all gave similar trends but neocortex ratio (ratio of neocortex size to that of the rest of the brain) gave the clearest effects. Measures of environmental complexity—range area, day journey length, and the amount of fruit in the diet—were found to be unrelated to neocortex ratio when body size effects were removed. In contrast, average species group size correlated with measures of neocortical enlargement, supporting a predominantly social function of primate intellect. Dunbar proposes specifically that neocortical size limits the social complexity that an individual can cope with: social complexity increases with group size, so groups begin to fragment when their size increases past a complexity limit set by neocortical size. These analyses can be criticized on statistical grounds—for instance, neocortical ratio is correlated with body size, and using species as data points brings problems of phylogenetic independence—but the broad findings have been confirmed, using absolute neocortex volume and the volume of the rest of the brain as independent predictors, and the method of independent contrasts to avoid
phylogenetic bias (Barton 1996; Barton and Dunbar 1997; Barton and Harvey 2000). Neocortical expansion is clearly linked with increased social group size in primates (and indeed also in bats, carnivores, and cetaceans).

These correlations give little clue, however, as to precisely what social skills are possible, or enhanced, by the possession of a larger neocortex. The problem is, of course, the difficulty of finding a species-fair and widely applicable measure of social skill, but there are now some steps in this direction. The most detailed analyses so far have been applied to the use of deception by primates within their own social group. Social manipulation of affiliated conspecifics, avoiding the disruptive use of violence, lies at the heart of the Machiavellian intelligence hypothesis (Humphrey 1976; Byrne and Whiten 1988). Cases where an individual achieves its ends by successfully deceiving another have long fascinated primatologists (Goodall 1971; de Waal 1982, 1986; Byrne and Whiten 1985), which made it possible to assemble an extensive corpus of carefully documented records, spanning all major groups of primates (Byrne and Whiten 1990). Preliminary analyses showed that most acts of deception were carried out with little sign of intentional understanding but nevertheless served effectively to manipulate the visual attention of other individuals and thence their understanding of the situations (Whiten and Byrne 1988a, b). These visual perspective-taking abilities were confirmed experimentally, in chimpanzees, when researchers modified naturally occurring situations of food competition for laboratory testing (Hare et al. 2000, 2001); the field data indicates, however, that visual perspective-taking is widespread also in monkeys. A larger corpus confirmed that only in the great apes did any of the accounts seem most parsimoniously explained as a result of mental state attribution (Byrne and Whiten 1991, 1992); even in these cases, alternative, nonmentalistic possibilities are entirely possible (Byrne 1993; Povinelli and Vonk 2003). However, the frequency of use of deception was clearly not uniform across species, nor easily explained as a result of observer effort (Byrne and Whiten 1992). Byrne and Corp (2004) investigated whether neocortical specialization was involved. Correcting the raw frequencies of observed deception for observer effort by using the number of long-term field studies of habituated individuals over the period, and using independent contrasts to correct for phylogenetic bias, they found that neocortical enlargement—whether measured by neocortex ratio or absolute volume—strongly predicted the rate of use of deception. In contrast, neither the volume of the rest of the brain, nor even the species’ average group size had significant effects (note that the number of species for which data were available was smaller than in Dunbar and Barton’s analyses of group size, so the difference may relate to statistical power). It seems that the use of deception for social manipulation critically depends on neocortex size.
What cognitive processes are likely to be involved in using deception? Given the lack, in the vast majority of cases, of any sign of intentionally planned deceit (Byrne and Whiten 1992; Byrne 1997), the main attribute would seem to be rapid memory in social contexts: memory of who was present on which occasion, who did what to whom, and so forth (Byrne 1996a). This suggestion finds support from the finding that social grooming is correlated with group size in Old World primate species (Dunbar 1993): for grooming to be valuable as a social currency, a good memory of grooming debts to and from social companions is essential. (It is sometimes doubted that any nonhuman animal is able to remember and use specific, time-referenced episodic information, e.g., Suddendorf and Corballis 1997; but recent experiments have shown clear evidence in scrub jays, a food-storing bird, Clayton and Dickinson 1998.) Innovation of novelty has been shown, similarly, to vary with neocortical enlargement in primates (Reader and Laland 2001). It is less clear what may be minimally needed for the ability to innovate successfully, but an analysis of the innovations employed in primate deception found that most cases involved only generalization of familiar behavior to slightly novel contexts (Byrne 2003b).

13.6 The intelligence to understand the world

Learning and memory are certainly critical aspects of intelligence to the extent that they are efficient, rapid, and flexible. But there is more to human intelligence than quick learning and reliable memory: to quote one definition, “grasping the essentials in a situation and responding appropriately to them” (Heim 1970). Putting this in more cognitive terms, intelligence implies the ability to represent the processes, social or physical, which are going on around us, and to use those mental representations to plan actions that may later be put into effect. Can we use evidence from living primates to deduce when, and in which ancestral species, this sort of intelligence evolved? Getting evidence of such capacities without using the medium of language is tricky, but this is a topic of extensive current research and some progress has been made. With very few exceptions, the evidence comes from great apes, although the interpretation of negative evidence is always problematic. Outside the great ape clade, it is very possible that primate behavior is not based on representational, mental models; even for great apes, interpretations remain controversial and are likely to change both as a result of improved data and sharper theoretical analyses.

Signs of social understanding, which would include some theory of mind abilities, have long been reported from analyses of observational data; these
have included empathy and sympathy, intentional deception, and pedagogical teaching (de Waal 1982; Boesch 1991; Byrne and Whiten 1991). Until recently, however, the consensus of laboratory experimentation was that great apes entirely lacked any such ability (Povinelli and Eddy 1996; Tomasello and Call 1997; Heyes 1998; Tomasello 1998). However, a number of experimental results have now bought the experimental and observational data sets into closer alignment (for reviews, see Call 2001, and Tomasello et al. 2003). Using more naturalistic paradigms to examine great ape problem-solving, a series of experiments have shown aspects of intentional understanding. These capacities closely match those apparent in some of the most complex cases of great ape deception, mentioned above (see Byrne and Whiten 1991): the ability to respond appropriately to differences in intention (e.g., accidental versus deliberate, inability versus unwillingness), visual perspective (hidden to a competitor but in view to self, partial versus completely hidden), and other individuals’ knowledge (e.g., known to one competitor but not another).

Signs of technical understanding have long been claimed for apes, because of the evidence of tool use and tool manufacture in the chimpanzee (and one population of orangutan), but there has been little attempt to work out exactly what understanding is needed for these skills to be learnt. Indeed, extensive negative data have suggested that perhaps chimpanzees have no reliable understanding of the physics of objects (Povinelli 2000), although positive, conflicting evidence of understanding comes from studies of tamarins, small New World monkeys which do not use tools (Hauser 1997), which emphasizes the inherent unreliability of all negative data. Many animals use detached objects as tools (Beck 1980), the process of tool manufacture is usually quite simple (McGrew 1992), and the main argument that great ape tool use relies on representational understanding is that sometimes tools are prepared or selected in advance, out of sight of the place of use (Byrne 1998). Most studies of chimpanzee tool use emphasize product rather than process, and in fact the evidence for unusual abilities in manual skill learning is stronger for the case of plant processing than tool use (Byrne 2004). Circumventing the physical defenses of herbivorous plants leads to the use of complex processing (Corp and Byrne 2002a), and both mountain gorillas and chimpanzees have been found to employ hierarchically organized procedures consisting of several modules employed in series or as subroutines (Byrne 1999b; Byrne and Byrne 1993b; Byrne and Russon 1998; Byrne et al. 2001; Stokes and Byrne 2001; Corp and Byrne 2002b).

Learning by imitation links social and technical intelligence: by means of social learning, technical skills are acquired. Imitation, in the rich sense of learning new, useful behavioral routines directly from observation, is often thought to rely
on prior understanding of mental entities, e.g., “the child must imaginatively place herself in the circumstances of the adult and determine what is the purpose of the behavior and how one goes about accomplishing that purpose” (Tomasello et al. 1993; my italics). Acquisition of novel behavioral routines by observation has been strongly argued for mountain gorillas, orangutans, and chimpanzees (Russon and Galdikas 1993; Byrne and Russon 1998; Byrne 1999a, b; 2002; Byrne and Stokes 2002; Lonsdorf et al. 2004), variously on the basis of (a) resemblance in fine detail between behavior of mother and offspring, (b) the fact that disabled individuals learn group-typical manual processes rather than devising more efficient, idiosyncratic versions, and (c) the sheer improbability of highly specific and complex organizations of behavior developing so similarly in each individual without some learning by imitation. It is also usually presumed that the site-specific differences in chimpanzee behavior reflect socially learned traditions in which useful skills are passed on by imitative learning of the critical aspects of processes which are hard for any individual to discover on any reasonable timescale, simple “cultures” (Whiten et al. 1999; Whiten 2000; but see Byrne et al. 2004 for caveats).

A tidy picture is thereby painted, of great apes—and perhaps only great apes—able to understand other individuals’ behavior in terms of intentional properties, giving them the ability to learn novel technical procedures from watching others who already have skills, leading to the elaborate, group-specific traditions in skills of significant technical sophistication described in wild chimpanzees. Given the close relationship of chimpanzees to ourselves, these similarities to human intelligence would be interpreted as resulting from common descent from a shared Miocene ancestor, having the ability to mentally represent intentional and causal-functional aspects of complex processes.

However, this picture may not be all it seems, however. Over just the same 20-year period in which evidence of richly complex behavior in great apes, once dismissed by experimentalists, has become hard to doubt, a number of theoretical analyses have begun to question what the evidence implies about the mental processes involved. It is important to stress that this challenge is not a reworking of early behaviorist critiques; behaviorism was hamstrung by its insistence on not postulating mental processes as “intervening variables,” whereas that is no longer at issue. However, once it is allowed that great apes are equipped with powerful systems for detecting and extracting patterns of statistical regularity in the world around them—patterns that correspond to intentions, plans, and cause and effect relationships—the use of statistical regularities may be sufficient to explain what the apes do, without their having concepts that correspond to human ideas of mental states and causal relationships. Those may all be dependent on language, and function in other ways than as primary causes of behavior.
Consider the case of imitation. There is a body of evidence, sketched above, that suggests great apes are able to learn by imitation the organizational structures of complex skills: how actions are grouped into often-used modules, how modules may be incorporated as subroutines into an overall plan (and thus repeated or omitted, depending on particular conditions of a given task), how sequential ordering affects outcomes, and so on. This is called program-level imitation (Byrne and Byrne 1993a; Byrne 1995b; Byrne and Russon 1998), and involves the use of hierarchical organization to structure novel actions. But where do the programs come from: is it necessary to discern the demonstrator’s intentions, and the cause–effect of how actions achieve their results, to derive a useful program? Not necessarily: if a sufficiently large corpus of behavior can be observed, then recurring patterns among the natural variability of behavior can in principle reveal all the organization underlying skilled action that is necessary for program-level imitation (Byrne 1999a, 2002): “imitation without intentionality” is a real possibility. The process of extracting statistical regularities from the messiness of natural, goal-directed action has been termed behavior parsing (Byrne 2003a), and serves to reveal the underlying deep structure of behavior, but it does not depend on explicit representation of intentions and causes. Instead, the extracted structures of behavior link prior circumstances to resulting outcomes, so that if a particular outcome is desired then those circumstances can be sought and that structure of behavior applied (see also Byrne 1995a, b).

In principle, it is possible that behavior parsing might also underwrite other behavioral routines that appear to rely on understanding ignorance, knowledge, intentions, and dispositions of others. While nothing so specific as for imitation has been worked out to date, Povinelli and Vonk (2003) point out that claimed mental processes of chimpanzees are “suspiciously similar” to those of humans, and suggest that this may be because humans (alone) construe behavior in those terms. The implication is that the cognitive system of nonhuman great apes is adept at extracting and using complex patterns of behavioral action but does not represent these patterns in the form of attributions about the mental states of others.

None of these critics suggest that humans lack the powerful mental processes for extracting statistical regularities from behavioral observations: their point is that the way we describe behavior in everyday talk, in terms of goals, plans, and intentions, may be uniquely human. Indeed, even 8-month-old infants are able to extract statistical regularities from spoken strings of letters, after only a few hours of exposure to monotonously spoken letter strings built according to particular rules (Saffran et al. 1996). Mention of “everyday talk” may give the clue to the function of intentional–causal representation: rather than primary
cause of behavior, such representations may be valuable because they allow pedagogy, explanation, and deliberate retrospective misrepresentation of behavior, using the medium of spoken language (Byrne 2006). Heretical as it may sound, it is worth questioning whether much everyday human behavior relies on intentional–causal analyses of situations: the alternative is that the same, powerful mechanisms for automatically extracting and using statistical regularities (regularities that themselves result, of course, from underlying intentions and causal dependencies) allow us to function in a “fast and mindless” fashion, responding appropriately and efficiently without deep thought about underlying mechanisms (and see Bargh and Chartrand 1999). Perhaps only when in contemplative mood, when asked directly, or when trying to explain (away) our actions, do we invoke the machinery of causal–intentional representation.

13.7 Moving forward

Whatever is the final resolution of this dispute over interpretation, it is now clear that great apes, and apparently not monkeys or other primates, behave in a range of sophisticated and humanlike ways. Neither allometric scaling of brain size against body size, however, nor expressing neocortical enlargement as a function of total brain size, produce a clear monkey/ape split in brain volumes. On Dunbar’s neocortex ratio, for instance, Papio baboons are as large brained as gorillas. The only brain measure that sets all the great apes completely apart from the monkeys is raw weight of neocortex (or brain), and this differs widely among apes in ways unrelated to any known behavior (for instance, a gorilla’s neocortex is much larger than a chimpanzee’s). Possibly the sharp intellectual differences between monkeys and great apes may result from a change in type of information processing or neural organization within the brain, but if so we still have no clue as to what it might be. It may be time to set comparative analyses of brain parts aside, and return to behavior—but this time avoiding the pitfalls of expecting some task or battery of tasks to give a single, quantified measure of intelligence, or taking a single species as “representative” of a wide range of related species in a taxon (as the rhesus macaque was taken to represent The Monkey, and the common chimpanzee The Ape).

It is convenient to speak as if intelligence were unitary, as if it could be measured on a single scale, but that is perhaps unlikely. If we accept that intelligence is not a single “thing” but rather a mixed bag of devices and processes, endowments and aptitudes, that together produce behavior we see as “intelligent,”
then it may make sense to study separate facets of intelligence independently of each other. Following Rozin (1976), modern psychometric work recognizes several “intelligences”: social–empathic, technical–mathematical, and common-sense–practical (Sternberg and Kaufman 2002). Presumably, in comparing across species, the different types of intelligence are likely to be even more sharply defined, and many cognitive capacities may need to be distinguished—each of them contributing to an impression of intelligent action but originating in different evolutionary circumstances and often at different periods of evolutionary history. Some psychologists have made strong arguments that intelligence is composed not just of several separate skill packages, but is massively modular, with all aspects of human intelligence underwritten by situation-specific skills, just as the blades on a Swiss Army knife are tools for many specific purposes. On the modular view, the human brain is made up of many discrete information-processing systems, each module reflecting adaptation to a characteristic problem in the Pleistocene environment of our ancient ancestry (Barkow et al. 1992). Thus, most of our everyday cognitive weaknesses and errors result from a mismatch between Pleistocene adaptation and modern problem: an extension of the familiar view that our evolved mental capacities cannot always keep up with a rapidly changing modern world. Undoubtedly, human intelligence is limited in many ways, but to insist a priori that there cannot be any sort of cognitive skill that is useful in many circumstances seems overly constraining. The view taken here is that deciding on the limits and scope of all cognitive capacities in humans and other animals is an empirical task.

All the same, from a practical point of view it will be safest to begin with the heuristic assumption that the skills required for each different task are unrelated and build up a data base on how a wide range of different species tackle each task—rather than making a priori assumptions as to what a particular task depends on, which has so often misled researchers in the past. This enterprise will be enriched by widening the scope beyond the nonhuman primates, to examine cognitively sophisticated members of other taxa entirely, and already a start has been made: with parrots (Pepperberg 1999), cetaceans (Herman 1986; Rendell and Whitehead 2001), pigs (Held et al. 2000, 2001, 2002), canids (Miklosi et al. 2004), and corvids (Emery and Clayton 2004). With converging evidence from all these taxa, and a solid body of data on a significant range of nonhuman primates, it will be time to attack once more the persisting problems of studying intelligence: How many kinds of intelligence are there? Is there a place for “general intelligence” in explanations? How does brain size relate to advanced cognitive skill? and so on.
References

Binet A, Simon TH (1915) Method of measuring the development of the intelligence of young children. Chicago Medical Book Company, Chicago
Boesch C (1991) Teaching among wild chimpanzees. Anim Behav 41: 530–532
Byrne RW (1999a) Imitation without intentionality. Using string parsing to copy the organization of behaviour. Anim Cogn 2: 63–72
Byrne RW (2003a) Imitation as behaviour parsing. Philos Trans R Soc Lond (B) 358: 529–536
Byrne RW, Byrne JME (1993b) Complex leaf-gathering skills of mountain gorillas (Gorilla...


Heim A (1970) Intelligence and personality. Penguin books, Harmondswort Middlesex


Suddendorf T, Corballis MC (1997) Mental time travel and the evolution of the human
Wechsler D (1944) The measurement of adult intelligence. Williams and Wilkins, Baltimore
14 Chimpanzee Hunting Behavior

Nicholas E. Newton-Fisher

Abstract

The pursuit, capture and consumption of small- and medium-sized vertebrates, appears to be typical of all chimpanzee (Pan troglodytes) populations, although large variation exists. Red colobus monkeys (Piliocolobus sp.) appear to be the preferred prey but intensity and frequency of hunting varies from month to month and between populations. Hunting is a predominately male activity and is typically opportunistic, although there is some evidence of searching for prey. The degree of cooperation during hunting, as well as prey selection, varies between East and West African populations and may be related to the way the kill is divided: in West Africa, hunters often collaborate, with kills tending to be shared according to participation, whereas in East Africa, the kill is typically divided tactically by the male in possession of the carcass, trading meat with females in return for sex or with other males to strengthen alliances, and cooperation in hunting is more limited. The adaptive function of chimpanzee hunting is not well understood, although it appears that it may be both a means to acquire a nutritionally valuable commodity that can then be traded and as a means for males to display their prowess and reliability to one another.

14.1 Introduction

The pursuit, capture and consumption of small- and medium-sized vertebrates, appears to be typical of all chimpanzee (Pan troglodytes) populations. Such hunting behavior has aroused considerable interest among anthropologists since it was first reported (Goodall 1963). Hunting, the division of the kill, and the consumption of meat all play an important role in the lives of modern hunter-gather societies (Lee 1979; Kaplan and Hill 1985; Hawkes et al. 2001; Hawkes and Bird 2002), and factor in a number of hypotheses concerning human evolution (Washburn and Lancaster 1968; Isaac 1978; Hill 1982; Tooby and DeVore 1987; Stanford 1998; Stanford 2001). While early ideas such as “Man the Hunter” (Washburn and Lancaster 1968) have largely been discredited, hunting as a means
to acquire meat remains important in many modern scenarios (Domínguez-Rodrigo 2002; Hawkes and Bird 2002). Animal tissue has high calorific value relative to plant material, is rich in fat and protein, and contains essential amino acids (Milton 1999). It is therefore a valuable resource. The nonrandom sharing of meat has been proposed as an important selective force driving the evolution of intelligence (Stanford 2001), and the consumption of meat has been invoked as an important proximate factor enabling the evolution of larger brains in the Homo lineage (Aiello and Wheeler 1995).

Chimpanzees show large variation between populations in the choice of prey species, frequency of hunting, and the techniques employed. Understanding both how and why chimpanzees hunt is important for the framing of evolutionary hypotheses; chimpanzees provide our best evidence for the behavioral capabilities of early hominines (Domínguez-Rodrigo 2002). In this chapter, I will review chimpanzee hunting behavior and attempt to address current hypotheses concerning the reasons why chimpanzees hunt, drawing out both similarities and differences between populations in their hunting behavior.

14.2 Chimpanzee hunting

All populations of chimpanzees that have been studied show some evidence of hunting and consuming vertebrate prey. Such hunting has been systematically documented among the East African chimpanzees (*Pan troglodytes schweinfurthii*) of the Gombe (van Lawick-Goodall 1968; Teleki 1973; Busse 1977; Stanford 1998) and Mahale (Nishida et al. 1979; Takahata et al. 1984; Uehara 1997) National Parks in Tanzania, of the Kibale Forest National Park (Mitani and Watts 1999; Watts and Mitani 2002) in Uganda, and among the West African chimpanzees (*P. t. verus*) of the Taï National Park, Cote D’Ivoire (Boesch and Boesch 1989; Boesch and Boesch-Achermann 2000). Other reports of hunting by chimpanzees come from East African populations in the Budongo Forest, Uganda (Newton-Fisher et al. 2002), Kahuzi-Biega National Park, DRC (Basabose and Yamagiwa 1997), Kasakati, Tanzania (Kawabe 1966), Semliki, Uganda (Hunt and McGrew 2002), from central African populations (*P. t. troglodytes*) of Lopé, Gabon (Tutin and Fernandez 1993), Ndoki, Cameroon (Kuroda et al. 1996; Takenoshita 1996), and from West African populations of Mt. Assirik, Senegal (McGrew 1983; Hunt and McGrew 2002), Bossou, Guinea-Bissau (Sugiyama and Koman 1987), and Tenkere, Sierra Leone (Alp and Kitchener 1993).
14.2.1 Prey diversity

Across populations, prey diversity is high with at least 40 species of vertebrates targeted. Chimpanzees are known to hunt a variety of primate species, as well as ungulates, rodents, birds, lizards, and frogs, and show a clear focus on mammalian prey (Table 14.1). Some chimpanzee populations appear to have a diverse range of prey, whereas others are more specialized. The Mahale chimpanzees, for instance, are known to hunt at least 17 species of mammals, while in Taï, chimpanzees hunt only 7 of the 15 sympatric mammal species, all primates (Boesch and Boesch-Achermann 2000; Boesch et al. 2002). Prey are typically small, up to a maximum of around 20 kg, the weight of an adult male black and white colobus monkey (Colobus guereza) (Kingdon 1997) or a part-grown bushpig (Potamochoerus porcus), but often much smaller (Goodall 1986).

14.2.2 Prey specialization

Monkeys, in particular colobus monkeys, appear to be the main prey of chimpanzees wherever the species are sympatric. Red colobus (Piliocolobus tephrosceles in East Africa, Pilicocolobus badius in West Africa) are the primary prey for many populations of chimpanzees, with black and white colobus (Colobus guereza in East Africa, Colobus polykomos in West Africa) as a secondary target.

The degree to which chimpanzees specialize on monkeys to the exclusion of other prey species varies between populations. In the Tai Forest, the chimpanzees show a notably strong specialization. Between 1984 and 1995, 93% of all prey were colobus monkeys: 80% black and white colobus (Colobus polykomos) and 13% red colobus (Piliocolobus badius) (Boesch and Boesch-Achermann 2000). A similar specialization is apparent among the Ngogo chimpanzees of the Kibale Forest, where between 1995 and 2000, 92.5% of all prey were colobus monkeys: 87.8% red colobus (Piliocolobus tephrosceles) and 4.7% black and white colobus (Colobus guereza). At Gombe, the specialization is less extreme but still noticeable: red colobus (there are no black and white colobus at this site) constituted 59% of the chimpanzees’ prey between 1970 and 1975, 66% between 1976 and 1981, and 84.5% between 1990 and 1995 (Goodall 1986; Stanford 1998).

In contrast, red colobus constituted only 53% of all prey for the Mahale chimpanzees (Nishida et al. 1992) and black and white colobus (Colobus guereza) were 43.8% of all prey for the Sonso chimpanzees in the Budongo Forest (Newton-Fisher et al. 2002), a location where there are no red colobus. These
<table>
<thead>
<tr>
<th>Table 14.1</th>
<th>Diversity of mammalian prey hunted by chimpanzees across Africa</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Species</strong></td>
<td><strong>Budongo</strong></td>
</tr>
<tr>
<td>Red colobus</td>
<td></td>
</tr>
<tr>
<td><em>Piliocolobus badius</em></td>
<td></td>
</tr>
<tr>
<td><em>Piliocolobus tephrosceles</em></td>
<td>✓</td>
</tr>
<tr>
<td>Black and white colobus</td>
<td>✓</td>
</tr>
<tr>
<td><em>Colobus guereza</em></td>
<td>✓</td>
</tr>
<tr>
<td><em>Colobus polykomos</em></td>
<td>✓</td>
</tr>
<tr>
<td><em>Colobus satanus</em></td>
<td>✓</td>
</tr>
<tr>
<td>Olive colobus</td>
<td>✓</td>
</tr>
<tr>
<td><em>Procolobus verus</em></td>
<td>✓</td>
</tr>
<tr>
<td>Gray-cheeked mangabey</td>
<td>✓</td>
</tr>
<tr>
<td><em>Lophocebus albigena</em></td>
<td>✓</td>
</tr>
<tr>
<td>Sooty Mangabey</td>
<td>✓</td>
</tr>
<tr>
<td><em>Cercocebus atys</em></td>
<td>✓</td>
</tr>
<tr>
<td>Olive baboon</td>
<td>✓</td>
</tr>
<tr>
<td><em>Papio anubis</em></td>
<td>✓</td>
</tr>
<tr>
<td>Yellow baboon</td>
<td>✓</td>
</tr>
<tr>
<td><em>Papio cynocephalus</em></td>
<td>✓</td>
</tr>
<tr>
<td>Vervet monkey</td>
<td>✓</td>
</tr>
<tr>
<td><em>Cercopithecus aethiops</em></td>
<td>✓</td>
</tr>
<tr>
<td>Red tail monkey</td>
<td>✓</td>
</tr>
<tr>
<td><em>Cercopithecus ascanius</em></td>
<td>✓</td>
</tr>
<tr>
<td>Campbell’s monkey</td>
<td>✓</td>
</tr>
<tr>
<td><em>Cercopithecus campelli</em></td>
<td>✓</td>
</tr>
</tbody>
</table>

**Chimpanzee hunting behavior**
<table>
<thead>
<tr>
<th>Animal Type</th>
<th>Species Name</th>
<th>Hunting Behavior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diana monkey</td>
<td>Cercopithecus diana</td>
<td>✓</td>
</tr>
<tr>
<td>L’Hoest’s monkey</td>
<td>Cercopithecus l’hoesti</td>
<td>✓</td>
</tr>
<tr>
<td>Blue monkey</td>
<td>Cercopithecus mitis</td>
<td>✓</td>
</tr>
<tr>
<td>Mona monkey</td>
<td>Cercopithecus mona</td>
<td>✓</td>
</tr>
<tr>
<td>Lesser spot-nosed monkey</td>
<td>Cercopithecus petaurista</td>
<td>✓</td>
</tr>
<tr>
<td>Crowned monkey</td>
<td>Cercopithecus pogonias</td>
<td>✓</td>
</tr>
<tr>
<td>Bushbaby</td>
<td>Galago sp.</td>
<td>✓</td>
</tr>
<tr>
<td>Potto</td>
<td>Perodicticus potto</td>
<td>✓</td>
</tr>
<tr>
<td>Chimpanzee</td>
<td>Pan troglodytes</td>
<td>✓</td>
</tr>
<tr>
<td>Ungulates</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forest duiker</td>
<td>Cephalophus sp.</td>
<td>✓</td>
</tr>
<tr>
<td>Blue duiker</td>
<td>Cephalohus moniticola</td>
<td>✓</td>
</tr>
<tr>
<td>Bushbuck</td>
<td>Tragelophus scriptus</td>
<td>✓</td>
</tr>
<tr>
<td>Suni</td>
<td>Neotragus moschatus</td>
<td></td>
</tr>
<tr>
<td>Bushpig</td>
<td>Potamocherus porcus</td>
<td>✓</td>
</tr>
<tr>
<td>Warthog</td>
<td>Phacochoerus aethiopicus</td>
<td>✓</td>
</tr>
</tbody>
</table>
Table 14.1 (continued)

<table>
<thead>
<tr>
<th>Others</th>
<th>Pan troglodytes schweinfurthii</th>
<th>P. t. troglodytes</th>
<th>P. t. verus</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Budongo</td>
<td>Gambe</td>
<td>Kahuzi</td>
</tr>
<tr>
<td>Giant elephant shrew</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Rhynchocyon</em> sp.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yellow-spotted hyrax</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Heterohyrax brucei</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White-tailed mongoose</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Ichneumia albicauda</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Civit</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Civettictis civetta</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rodents (various spp.)</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Pangolin</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Populations (study sites) of chimpanzees are arranged by subspecies. A ✓ indicates that the species has been recorded being killed or consumed. Not all species of prey are present at all sites.
two populations appear to differ from the others in that the chimpanzees also prey upon small ungulates, particularly blue duiker (*Cephalophus monticola*), to an appreciable degree: 34% of all prey in Mahale (Nishida et al. 1992) and 25% of all prey in Budongo (Newton-Fisher et al. 2002). Data from Budongo are sparse, but recent observations support the idea that these chimpanzees do not demonstrate the extreme prey specialization seen in Taï and Ngogo (Newton-Fisher unpublished data). Forest ungulates, particularly duiker and bushpig, are in fact hunted by all the East African chimpanzee populations studied (Gombe: Goodall 1986; Mahale: Nishida et al. 1992; Budongo: Newton-Fisher et al. 2002; Kibale: H. Sherrow personal communication), but do not appear to be regarded as prey by West African chimpanzees (Uehara 1997; Boesch and Boesch-Achermann 2000). More research is needed on chimpanzee predation on ungulates.

Chimpanzee populations also appear to differ in their choice of the age and sex of prey. For the Taï chimpanzees, half of their colobus monkey prey were adults, mostly females (Boesch and Boesch-Achermann 2000). This is in contrast to chimpanzees at Mahale and Gombe, where the vast majority of colobus prey were juveniles and infants (Goodall 1986; Uehara 1997) and some chimpanzee hunters target very young colobus monkeys, snatching them from their mothers (Stanford 1998). There is less information on the age and sex of non-colobus prey. Among the ungulates, bushbuck are targeted only as infants (fawns), as typically are bushpig (piglets) (Goodall 1986). Age and sex estimates of duiker kills are more difficult to obtain, given that the prey is rapidly torn apart and consumed entirely by the chimpanzees, however it seems clear that chimpanzees are quite capable of killing adult blue duiker (personal observations).

### 14.2.3 Sex bias in hunting

Hunting by chimpanzees is a predominately male activity. Among the chimpanzees of the Ngogo community in the Kibale Forest, adult or adolescent males made 98.8% of all kills recorded between 1995 and 2000 (Watts and Mitani 2002). In two decades of data from Gombe, adult males were responsible for 91.5% of all kills (Stanford 1998). Female chimpanzees will and do hunt, however. Data from Gombe for 1977–1979 showed that females joined an average of 26% (median: 25%, range: 0%–67%) of red colobus hunts for which they were present, and those females that were more likely to join males in a hunt were also more likely to hunt when apart from the males (Goodall 1986). One female, Gigi, contributed 4% of the total kills (Stanford 1998). Any kills that females made during a mixed sex hunt were likely to be taken by males (Goodall 1986), which may in part explain female unwillingness to hunt when with males. Females may prey more
on ungulates (Uehara 1997), but quantitative data are difficult to collect, in part due to the nature of ungulate hunting.

14.3 Hunting frequency

Detecting hunting in a chimpanzee population can be problematic, particularly, if the chimpanzees are poorly habituated. Typically in this situation, hunting is rarely if ever seen, and studies rely on finding animal remains, skin or bone, in chimpanzee feces (McGrew 1992). Unfortunately, sampling feces does not appear to be a reliable indicator of the occurrence of hunting; while the presence of remains can confirm that consumption does occur, little can be said about the frequency of consumption (cf. Uehara 1997). Long-term observations of habituated chimpanzees in the Taï Forest have revealed a pattern of frequent hunting and consumption that is not mirrored in the pattern of prey remains found in fecal samples (Boesch and Boesch-Achermann 2000). Further, fecal sampling can say nothing about the number of hunting attempts that fail to secure prey, the division of the prey once obtained, or the relative importance of scavenging as a method of acquiring meat. Similar problems may also occur when hunting is actually rare, or when prey species are alerted or scared away by the presence of humans accompanying the chimpanzees, although in some cases chimpanzees may exploit their prey’s fear of humans to increase hunting success (Goodall 1986; Boesch 1994).

14.3.1 Predation pressure

While in some populations chimpanzees appear to hunt only rarely, in other populations they are significant predators who may hunt at levels that appear to be unsustainable (Goodall 1986; Wrangham and van Zinnicq Bergmann Riss 1990). Estimates for Gombe suggest anything from 8% to 42% of the colobus population can be killed each year, with the number varying from year to year: 8%–13% (1973–1974: Busse 1977), 41.6% (1972–1975: Wrangham and van Zinnicq Bergmann Riss 1990), 16.8%–32.9% (1982–1991: Stanford et al. 1994), while at Taï during the 1980s, the figure was between 3% and 8% (Boesch and Boesch-Achermann 2000). In contrast, the Mahale chimpanzees were estimated to kill only around 1% of the red colobus population each year during the 1980s (Boesch et al. 2002). Basabose and Yamagiwa (1997) estimate that the chimpanzees of Kahuzi-Biega kill 11%–18% of the Cercopithecus monkey population each year (predominately Cercopithecus mitis but also Cercopithecus l’hoesti). Hunting
of ungulates may also impose high levels of mortality. Wrangham and van Zinnicq Bergmann Riss (1990) estimated chimpanzee-imposed mortality on bushbuck at 27% (although this figure includes bushbuck fawns killed by baboons and subsequently stolen by chimpanzees) and on bushpig at 7%, for populations in the Gombe National Park between 1972 and 1975.

These estimates, both for primates and ungulates, are based on comparing the number of kills with the population density of prey within the chimpanzee community’s home range. There is potential for error in the estimates of each of these variables. If, for example, home range is overestimated (cf. Newton-Fisher 2004), then predation pressure will be underestimated, while underestimating the number of potential prey will inflate the estimate of predation pressure (Wrangham and van Zinnicq Bergmann Riss 1990).

14.3.2 Variation in hunting frequency

Such estimates of hunting frequency and predation pressure, however, disguise wide variation. Within a single community, the total number of hunts can vary from month to month and year to year. Across populations, chimpanzees appear to have hunting “seasons,” during which the number of kills increases either as a result of more hunting or more successful hunting, or both. For the chimpanzees at Gombe, Mahale, and Taï, this hunting season falls toward the end of the year, peaking in September and October. At Gombe, this corresponds to the later part of the dry season (Stanford 1998). At Mahale, the peak is slightly later, reaching into November and appears to coincide with the end of the dry season and the first rains of the wet season (Takahata et al. 1984). Preliminary work at Budongo suggested a dry season (December to February) peak in hunting activity (Newton-Fisher et al. 2002), but subsequent work has failed to confirm this idea (Newton-Fisher unpublished data).

The hunting behavior of the Ngogo chimpanzees does not appear to correspond to timing of rainfall, but hunting seasons instead occur during periods of fruit abundance (Watts and Mitani 2002) that are not correlated with rainfall (Mitani et al. 2002). Similarly, the hunting season at Mahale occurs when more fruit is available (Uehara 1997). At Taï, the hunting season runs from mid-August to mid-November, between periods of low and high fruit abundance and ending when chimpanzees switch to highly calorific Coula edulis nuts from which they gain sugar, protein, and fat. The peak in hunting is also in September and October, but this is during the wet season at the time of greatest rainfall (Boesch and Boesch-Achermann 2000).
In addition to these seasonal changes, hunting frequency within a single community varies between years. Such changes might be related to changes in the abundance of prey species or the number of chimpanzees who might hunt. A comparison of hunting success for Mahale chimpanzees between the 1980s and early to mid 1990s showed a threefold increase in the percentage of the red colobus population killed by the chimpanzees, rising from around 1% to at least 3% of the population per year (Boesch et al. 2002). This seemed to accompany an expansion in the red colobus population. Hunting success then fell in the later part of the 1990s, following a drop in the number of chimpanzees in the study community (Boesch et al. 2002). A similar drop in hunting was seen following a decrease in the number of adult males in the study community in the Tai Forest (Boesch and Boesch-Achermann 2000). Chimpanzees may also experience greater hunting success when individuals with a flair for hunting are present. These individuals demonstrate both a high willingness to hunt and a consistently high probability of success (Goodall 1986; Stanford 1998; Boesch and Boesch-Achermann 2000).

### 14.3.3 Hunting binges

A further source of variation in hunting frequency within a community is the occurrence of hunting “crazes” (van Lawick-Goodall 1968) or “binges” (Stanford 1998). These are periods during which the chimpanzees hunt “almost daily”: more than three hunts in a 7-day period, with chimpanzees appearing to hunt on contact with prey (Stanford 1998). In the Kasekela (Gombe) community, 23 binges were recorded between 1990 and 1995. The longest of these lasted 74 days and consisted of 38 observed hunts and at least 76 kills, all red colobus. Correcting the number of kills for observation time suggests that over 100 colobus monkeys were killed during this 74-day period (Stanford 1998). The Ngogo chimpanzees went on a 57-day hunting binge in 1998, during which they hunted 22 times, killing 69 red colobus, 1 mangabey (*Lophocebus albigena*), and 1 red duiker (*Cephalophus* sp.). Only 4 of these 22 hunts were unsuccessful, including two attempts to hunt black and white colobus (*Colobus guereza*). This and other hunting binges at Ngogo coincided with major fruit crops, and most hunting occurred when large parties of males were traveling together (Watts and Mitani 2002). Large parties with high numbers of males also seem to be linked to hunting binges at Gombe (Stanford 1998). Large numbers of chimpanzees traveling together suggests that fruit is particularly abundant, and so hunting binges at Gombe may also be linked to periods of food abundance.
14.4 How do chimpanzees hunt?

Many of chimpanzee hunts are opportunistic, in that chimpanzees appear to decide to hunt after encountering prey during the course of normal foraging activities and travel around the home range. This seems to be the typical pattern at Gombe (Goodall 1986; Stanford 1998) and at Ngogo (Mitani and Watts 2001). In contrast, chimpanzees in Taï appear to search actively for prey, listing for the vocalizations of either colobus monkeys or of Diana monkeys (*Cercopithecus diana*) with whom the colobus are frequently associated (Boesch 1994; Boesch and Boesch-Achermann 2000).

Chimpanzees hunt without the use of tools or weapons, although there are a few reports of rocks or branches being hurled, possibly in an attempt to panic defensive formations of adults (Goodall 1986), and tools (sticks and leaves) are sometimes used to aid in the processing of the carcass (McGrew 1992).

During a hunt, prey are typically chased, seized, and then killed either by a bite, disembowelment or by being torn apart (Goodall 1986). Hunts may yield single or multiple kills. Between 1973 and 1981, Gombe chimpanzees made multiple kills in 37.5% of colobus hunts; most of these were two kills per hunt. A typical colobus hunt at Gombe will produce two (Watts and Mitani 2002) or three (Stanford 1998) kills and at Ngogo, four kills (Mitani and Watts 2001). Single kills seem to be more usual for Taï chimpanzee hunts (Stanford 1998), although these will typically be of adult monkeys (mean number of kills per successful hunt: 1.2; Watts and Mitani 2002).

Chimpanzees will hunt alone and in the company of others. Solitary hunts occur rarely at Taï (16% of hunts: Boesch and Boesch-Achermann 2000) and Mahale (28% of hunts: Takahata et al. 1984; Uehara et al. 1992), while they are more common at Gombe (64% of hunts: Busse 1978; Teleki 1973) where the chimpanzees appear to be highly effective solo hunters. Boesch (Boesch 1994) calculated that Gombe chimpanzees had a success rate of 50% when hunting alone, capturing an average of 1.6 kg of prey within 7 min of hunting. In contrast, his estimate for the success rate of lone hunters at Taï was only 13%, yielding an average of 9.5 kg of prey within 39 min of hunting. The forest canopy is lower and more broken at Gombe than it is at Taï, which may make it easier for lone chimpanzees to isolate colobus monkeys and so allow them to capture and kill their prey more often and more quickly (Boesch 1994). This idea is supported by data from Ngogo, where chimpanzees were more likely to hunt when encountering prey in, or close to, areas with broken or no tree canopy than when in primary forest (Watts and Mitani 2002).

Group hunts are often a case of individual chimpanzees making their own efforts in a collective setting, perhaps exploiting the panic in the prey produced by
the presence of multiple hunters, and reacting to the actions of other chimpanzees. Collaborative hunting, where males taking particular roles such as “drivers” and “blockers” (Boesch and Boesch 1989), appears to be the primary form of hunting among the Täi chimpanzees (77%: Boesch and Boesch-Achermann 2000) but is rare among the East African chimpanzees (Boesch 1994; Stanford 1998; Boesch and Boesch-Achermann 2000; Watts and Mitani 2002). A division of roles between those that pursue the prey and those that wait on the ground to capture monkeys that fall from the canopy is, however, fairly common among East African chimpanzees.

14.4.1 Cooperative hunting?

There is little consensus over the extent to which chimpanzee hunting is cooperative. To the extent that chimpanzees take different roles, and are responsive to one another’s behavior during a hunt, there is good evidence for social cooperation, but to demonstrate that chimpanzee hunting is functionally cooperative, individuals need to do better when hunting as a group. Thus, if cooperation occurs, hunting attempts should be more successful when more individuals take part, or at least certain number of hunters should be more successful than solitary hunters. At Gombe, Ngogo, and Täi, the probability of killing prey during a red colobus hunt increases with the number of hunters present, but this appears to be a simple effect of more hands grabbing at the monkeys; there does not appear to be an additional effect from males working together (Stanford 1998). At none of these sites does the mass of prey per individual hunting correlate with the number of males hunting (Stanford 1998; Watts and Mitani 2002), although Gombe chimpanzees do gain a higher return (greater mass of prey per hunter) when more than seven are hunting together (Stanford 1998). Many male chimpanzees hunting together may be able to overwhelm the defensive strategies of the red colobus, and reduce the opportunities for panicked monkeys to escape. Among Täi chimpanzees, the number of hunters is strongly correlated with the mass of prey caught because the likelihood of capturing an adult monkey increases but gains per hunter peak at four males (Stanford 1998), presumably because most hunts terminate after the first kill.

The hunting of ungulates is less well described. Bushpigs are probably the most difficult of ungulate prey. Chimpanzees are wary, if not fearful, of the adults and they retreat to the trees in the face of aggression by adult pigs. At Gombe, chimpanzees have been described using stealth to seize piglets before the adults are alerted to their presence, and also of using aggressive displays to panic the adults, capturing piglets either in the confusion or if abandoned by adults that run off (Goodall 1986).
Bushbuck fawns hide in dense cover as an antipredator strategy, while adults typically freeze or flee. Chimpanzees search for hiding fawns when their attention is drawn to particular areas by the presence of adult bushbuck or possibly auditory or olfactory cues. A captured fawn’s mother may be aggressive toward chimpanzees, but this is difficult to determine as human presence causes them to flee (Goodall 1986). Duiker captures are typically opportunistic, with chimpanzees seizing them if they come within reach. Chimpanzees sometimes show interest in duiker vocalizations (personal observations) but the extent to which they search for duiker is unclear.

14.4.2 Scavenging

Chimpanzees are reluctant scavengers. Only a handful of reports exist of scavenging, acquiring meat from an abandoned carcass. Most of these observations concern the seizing of fresh kills from other predators, a behavior often labeled piracy (Goodall 1986; Uehara 1997; Stanford 1998), although plundering, the forcible stealing of goods, might be more appropriate term. At Gombe, chimpanzees have been recorded seizing fresh kills from baboons (Morris and Goodall 1977; Goodall 1986), and at Budongo, the body of infant blue monkey (Cercopithecus mitis) was stolen from the adult blue monkey who killed it (Newton-Fisher et al. 2002). Boesch and Boesch-Achermann (2000) report three instances of Tai chimpanzees robbing eagles of red colobus captures while the monkeys were still alive, and a further four instances of chimpanzees eating the kills of eagles; presumably these were recently killed monkeys, but this information is not reported. Given that chimpanzees are quite willing, if they can steal or beg part of the carcass, to eat prey that chimpanzees other than themselves have killed, it is not surprising that they are similarly willing to take fresh kills from other species.

True scavenging, however, appears particularly rare. In 36 years of observation at Gombe, fewer than 20 instances have been recorded (Stanford 1998), and at least 9 of these (all red colobus) are likely, or known, to be previous chimpanzee kills (Goodall 1986). Similar low rates have been recorded at Mahale; seven cases in over 25 years of observation: six ungulates and one red-tailed monkey (Cercopithecus ascanius) (Hasegawa et al. 1983; Uehara 1997). Scavenging has not been reported from Tai: Boesch and Boesch-Achermann (2000) record ten encounters with fresh carcasses, none of which were eaten by the chimpanzees.

Most encounters with fresh carcasses result in apparent curiosity-type behaviors in the chimpanzees, with no indication that the chimpanzees regard these carcasses as a source of meat. Stanford (1998) reports an observation from Gombe of a juvenile male briefly chewing on 1- or 2-day-old colobus meat that
was ignored by the adults, and Muller et al. (1995) record a further observation, from the same community, of a party of chimpanzees encountering a dead bushbuck, presumed to be killed by a leopard. The chimpanzees showed strong curiosity over the carcass, even grooming it, and one female rolled around inside the eviscerated carcass, but they did not feed (Muller et al. 1995). In contrast, chimpanzees at Mahale did feed on the carcasses of two adult bushbuck thought to be the remains of leopard kills (Hasegawa et al. 1983).

14.5 Meat eating

All populations of chimpanzees subsist on a primarily frugivorous diet. Typically, fruit constitutes 60%–80% of the time spent feeding (Gombe: 63%, Wrangham 1977; Kibale: 79%, Wrangham et al. 1996; Budongo: 64.5%, Newton-Fisher 1999a). This is supplemented by leaves, primarily, as well as other plant materials. Even in communities that hunt frequently, such behavior constitutes a very small portion of the time spent foraging. Watts and Mitani (2002) recorded 131 predation episodes in 6 years at Ngogo (1.8 hunts per month) while Boesch and Boesch-Achermann (2000) recorded 413 hunts in a 12-year period at Taï (2.9 hunts per month). Nevertheless, chimpanzees do hunt, kill, and consume meat, and competition over the division of the kill can be high.

14.5.1 The value of meat

A single carcass can represent a sizable resource, despite substantial variation between species. *Colobus guereza* adults weigh up to 23 kg, although Ugandan populations may not reach this size, while the western black and white colobus (*Colobus polykomos*: adult male body weight: 8–12 kg) is smaller and similar in size to the eastern red colobus (*Piliocolobus tephrosceles*: adult male body weight up 13 kg). The western red colobus (*Piliocolobus badius*) are lighter, with an adult body weight of only 5–10 kg (Kingdon 1997).

Commonly, the entire animal is consumed, including bones and skin, and chimpanzees will compete for the smallest scraps. The chimpanzees will therefore be consuming not only the protein-rich flesh, but also fat, notably from the brain, and bone-marrow, together with minerals present in the blood and bones (Stanford 1996; Stanford 1998).

The quantity of meat, including the associated elements of the carcass, consumed by some individuals may be relatively significant. In good hunting years, the total amount of meat consumed may be more than double that...
consumed in poorer years. The 45 Gombe chimpanzees of the Kasakela community in 1992 consumed over 500 kg of red colobus meat, and their total meat consumption for the year was probably close to 700 kg. The previous year (1991), colobus meat consumption was less than 200 kg, and in 1988, this figure was less than 150 kg (Stanford 1998). Averaged over years, the level of consumption in the 1980s and 1990s seems similar to the estimate of 441 kg of meat per year for the same community in the 1970s (Wrangham and van Zinnicq Bergmann Riss 1990; Stanford 1998).

Boesch and Boesch-Achermann (2000) estimate that, averaged across the year, male Taï chimpanzees consume 186 g per day, while females consume 25 g per day. Their estimates for Gombe chimpanzees, similarly averaged, are 55 g per day for males and 7 g per day for females. These are similar to estimates made by Stanford (1998) of 70 g per day for males during peak hunting season and by Wrangham (1975) of 22 g averaged over males and females.

Meat therefore appears to be a valuable resource to chimpanzees, although there are exceptions. Captured prey may be only partly eaten before being discarded. In the Taï Forest, adult cercopithecine monkeys have been treated in this way (Boesch and Boesch-Achermann 2000). At Gombe, chimpanzees have been observed discarding captured adult red colobus in favor of pursing immature monkeys (Boesch 1994; Stanford 1998) and giving a carcass to another individual in order to hunt again (Goodall 1986). In the Budongo Forest, an adult male chimpanzee captured and killed an elephant shrew (Rhynchocyon sp.) but took only a single bite before discarding the carcass (Newton-Fisher unpublished data), and the bodies of infant chimpanzees killed by adults are sometimes only partially eaten before being handed on to another individual or discarded completely (Newton-Fisher 1999b). Furthermore, any kills made during a group hunt are typically divided in some way among some or all of the chimpanzees present.

14.5.2 Begging and food sharing

Following a kill there is commonly a degree of competition for the meat, the intensity of which reinforces the idea that chimpanzees desire and value meat. If the chimpanzee in possession of the carcass has companions, these individuals will attempt to acquire part of the carcass. More dominant individuals may attempt to steal the entire carcass for themselves. Others will sit around the possessor and beg for a share of the meat. Begging individuals seem to exert a lot of pressure both by their presence and by their harassing gestures and vocalizations. Chimpanzees unwilling to share will commonly move away from
the crowd of begging individuals, although they are likely to be followed. When harassed by one or two others, a chimpanzee may simply turn its back toward them to avoid them reaching toward the carcass.

Sharing of the carcass can be either an active or passive process. Most sharing is passive and ranges from an individual patiently scrounging the scraps that fall from a carcass as the possessor feeds, through harassment of the owner of carcass by gestures and vocalizations, to an individual who is not in possession of the kill taking a portion of carcass without the use of aggression. Active sharing is less common and involves the individual who possesses the carcass handing part, or all, of the carcass to another chimpanzee. There are a number of theories to explain why food should be shared, and the patterns of sharing observed. These including tolerated theft, reciprocity, kin selection, mutualism, buy-off, and harassment. As they apply to chimpanzees, these theories have been discussed elsewhere (de Waal 1989; Mitani and Watts 2001; Fruth and Hohmann 2002; Stevens 2004; Stevens and Gilby 2004).

Patterns of sharing appear to differ between West and East African chimpanzees. In the Taï Forest, West African chimpanzees tend to divide the kill among the individuals who participated in the hunt. Older and more dominant males gain a greater share of the meat, but hunters tend to receive more than nonhunters, even when socially subordinate. The amount of meat obtained by females is not dependent on participation in the hunt, but females will support hunters over nonhunters when there is competition (Boesch and Boesch-Achermann 2000). In East Africa, at Gombe (Stanford 1998), Mahale (Nishida and Hosaka 1996), Ngogo (Mitani and Watts 2001), and Budongo (Newton-Fisher unpublished data), chimpanzees use a different strategy for the division of the carcass. Males tend to monopolize the carcass and share only with particular adult females and other adult males, biasing their sharing toward alliance partners.

The sharing strategy employed by West African chimpanzees may oblige them to hunt adult monkeys. Collaborative group hunting appears necessary to increase hunting success and to reduce the time spent hunting in a habitat that favors escape by the prey but may only work if males are rewarded for participating in the hunt (Stanford 1998). Colobus monkeys are smaller in West Africa than they are to the East, which might make targeting juveniles unprofitable if the meat has to be shared among all hunters.

For East African chimpanzees, the larger body size of the colobus monkeys may pose a greater hazard and East African chimpanzees show greater fear of adult colobus monkeys than do those in West Africa. Adult colobus monkeys can successfully threaten and rout chimpanzees, chasing them from trees on occasion (Nishida et al. 1979; Goodall 1986; Boesch and Boesch 1989). Given that the
strategy adopted by East African chimpanzees of targeting juvenile and infant chimpanzees appears to be profitable (Boesch 1994), the additional costs of targeting adult monkeys together with the more Machiavellian sharing strategy may make hunting adult monkeys a less attractive and unnecessary option.

14.6 Why do chimpanzees hunt?

Meat is a nutritious food and it seems likely that chimpanzees who eat meat, particularly those males who consume relatively large quantities, gain nutritional benefits. Nevertheless, the frequency of consumption varies from individual to individual, month to month, year to year, and between populations. Chimpanzees are omnivores, carnivory does not appear to be critical for survival or reproduction, and various theories have been advanced to explain the occurrence of their hunting behavior.

14.6.1 Hunting for nutrition

Early views of chimpanzee hunting favored the view that hunting was driven by nutritional demands. Teleki (1973) proposed that Gombe chimpanzees may hunt to compensate for nutritional shortfalls, given the strong seasonality at this site. The body weights of Gombe chimpanzees are lower during the dry season (Williams et al. 2002), which may to be the consequence of low-food availability, and hunting at Gombe is more pronounced during the dry season than it is during the wet season (Stanford 1998). A nutritional perspective was also emphasized by Wrangham (1975), with a similar view emerging from research at Mahale (Takahata et al. 1984).

14.6.2 Hunting for trade goods

The nutritional content of meat and associated tissues, together with the fact that it is both divisible and portable, means that each portion has an inherent value and can be either consumed or given to another individual. Meat can thus be considered a commodity that can be traded with other individuals for other goods or services, which for chimpanzees are likely to be biases in future social interactions such as support in agonistic confrontation or increased levels of grooming. Such a “biological markets” (Noe and Hammerstein 1995) perspective is implicit in two further theories concerning chimpanzee hunting, both of which
see an adaptation in the nonrandom sharing of kills, with hunting simply the most common behavior that provides access to carcasses.

The first of these, labeled “meat-for-sex” by Mitani and Watts (2001), was also proposed by Teleki (1973). He noted that cycling females with conspicuous ano-genital swellings tended to receive meat from adult males more frequently than did females without these sexual swellings, and suggested that males shared meat with females in exchange for sexual access. Swollen females are attractive to males (Dixson 1998) as the swellings generally indicate approaching ovulation, although females will also show swellings when pregnant (Wallis and Lemmon 1986). Supporting evidence for this hypotheses was provided by Stanford (1998) who found that, at Gombe, the presence of a swollen female in a party of chimpanzees was the best predictor of a hunt occurring when encountering a group of red colobus, and reported five observations of females begging for meat from males, and only being given part of the kill after copulating.

The other theory that involves using prey as a trade good is the “male-social-bonding” hypothesis. Nishida (Nishida et al. 1992; Nishida and Hosaka 1996) provided data to support the idea that males trade meat with other males in order to develop and maintain the alliances that play an important role in male–male competition for status. Mitani and Watts (2001) showed that, at least for the Ngogo chimpanzees, while the presence of a female with a sexual swelling was a significant predictor of the decision to hunt, this was an artifact of the relationship between the presence of such females and the number of adult males, and that it was the number of adult males alone that predicted hunting. They also showed that males shared reciprocally, at least when considering all pairs of males simultaneously, and that there was a positive association between sharing of carcasses and support in agonistic coalitions (Mitani and Watts 2001; Watts and Mitani 2002).

14.6.3 Hunting to assess reliability

Male chimpanzees vary in their hunting ability, as demonstrated by the proportion of hunts that they join, the number of kills that they make, and their success at hunting alone (Stanford et al. 1994; Stanford 1998; Boesch and Boesch-Achermann 2000; Watts and Mitani 2002). Among the Ngogo chimpanzees, and potentially elsewhere, good hunters are also more frequent members of the territorial patrols that monitor and probe boundaries with neighboring communities. Furthermore, males that hunt together patrol together, and the frequency of joint patrolling is correlated with the frequency with which males form coalitions and the amount of grooming between them (Watts and Mitani 2001;
Watts and Mitani 2002). This leads to the hypothesis that hunting itself may have a function that is independent from acquiring meat: it demonstrates risk-taking and allows males to assess the reliability of others when faced with danger (Watts and Mitani 2001). This is essentially a refinement of Kortlandt’s (1972) “hunting-to-display-social-prowess” hypothesis. Given the risks associated with patrolling and intercommunity encounters (Goodall et al. 1979; Boesch and Boesch-Achermann 2000; Muller 2002), such assessments may be important for male chimpanzees. This hypothesis may be important in relation to the “showoff” hypothesis proposed to explain hunting behavior in human males (Hawkes 1991; Hawkes and Bird 2002).

14.6.4 So who do chimpanzees hunt?

This question, which addresses the adaptive value of hunting, remains to be answered. It is only recently that quantitative analyses comparing the different hypotheses have been undertaken (Mitani and Watts 2001) and, while there are efforts to draw together results from different populations (Uehara 1997; Boesch et al. 2002), systematic analyses have yet to be applied across populations.

The possibility that chimpanzees achieve nutritional benefits directly from hunting cannot be easily dismissed. The necessary nutritional studies quantifying chimpanzee diet have not been conducted, and for either of the trade-goods hypotheses to operate, there must be a nutritional gain to the individuals who receive and consume parts of the carcass. If there were not, the carcass would hold no value and could not be traded. The particular hypothesis that chimpanzees switch to hunting to compensate for nutritional shortfalls finds little support from the only direct test: Mitani and Watts (2001) found that Ngogo chimpanzees hunted more frequently as fruit became more abundant, suggesting that chimpanzees are more likely to hunt when they have enough surplus energy. This makes sense if hunting is energetically costly, and individuals risk not gaining enough meat following division of the kill to offset such costs.

Whether this relationship between food abundance and frequent hunting applies to all populations of chimpanzees remains to be determined, but hunting seasons do coincide with fruit abundance in Mahale (Takahata et al. 1984; Uehara 1997), although apparently not at Taï (Boesch and Boesch-Achermann 2000) where chimpanzees may gain shares of the kill that depend on their participation in hunting (Boesch 1994). If the Taï chimpanzees capture and kill a sufficiently large prey in each hunt, and if they can rely on this system of dividing the meat, then net nutritional gains would accrue to all participants.
At Gombe, the relationship between seasonal changes in body weight and condition and hunting remains intriguing, although Stanford (1998) concludes that the nutritional shortfall hypothesis is unlikely to explain hunting in this population. Prey are typically small, often immature monkeys, and division of the kill is unequal. As a result, hunters cannot rely on recouping lost energy.

It appears that East African chimpanzees, at least, hunt to gain meat but not to compensate for nutritional shortfalls. This meat is used as a commodity and traded with others. While the “meat-for-sex” and “male-social-bonding” hypotheses are commonly presented as alternatives (Mitani and Watts 2001; Watts and Mitani 2002), they may be alternate, context-dependent, outcomes of the same social strategy. The “meat-as-commodity” hypothesis proposes that chimpanzees hunt to gain possession of a commodity (part or all of an animal carcass) which has economic value within chimpanzee society (Stanford 1998). They can then trade this to further whatever proximate goals are most pressing, providing meat to allies when they have need of them or to females in an effort to coerce their mating behavior.

At Gombe, the presence of a female with a sexual swelling was the best predictor of hunting and this remained true when the effect of number of males in the party (also a predictor of hunting) was removed. Males shared food with swollen females, and in some cases shared only after copulating (Stanford 1998). At Ngogo, by contrast, the number of males was the best predictor of hunting and the presence of swollen females was not a significant predictor once the effect of the number of males was removed. Males of this community did preferentially share meat with swollen females, but did not copulate with those females after sharing at a level above chance, and did not gain a larger share of matings if they shared with a female, comparing female cycles in which the male shared with those in which he did not (Watts and Mitani 2002). It is not clear, however, whether Ngogo males shared before or after copulating; Gombe males were described as withholding meat until the female copulated.

If the “meat-as-commodity” hypothesis holds, then these apparently contradictory findings on the reasons for chimpanzee hunting can be reconciled. If the immediate proximate benefits for the males possessing the meat differed between the two communities, males would be predicted to use their “trade goods” in different ways. Important factors are likely to include the number of males in the community and the intensity of within-community competition for status, which will translate into the opportunities and necessity for alliance formation, respectively. The two communities differ markedly in this respect. The Gombe community in the early 1990s numbered 45 individuals, with 8 adult males (Stanford 1998), whereas the Ngogo community (in the late 1990s) numbered in excess of 140 individuals, including 24 adult males (Mitani and Watts 2001). Other
important factors may include the degree to which males can coerce female mating behavior through other means, and differences in female mating strategies. For example, if Gombe females are nutritionally stressed during the dry season, they may be prepared to compromise a promiscuous mating strategy by providing sex in return for even the smallest scraps of food. McGrew (1992) showed that those females who were more successful at getting meat had greater numbers of surviving offspring. Ngogo females may not be under the same pressure, and they may place less value in male-controlled meat. Testing whether “meat-for-sex” and “male-social-bonding” are different proximate outcomes of the same adaptation awaits future research.

Nutritional gain appears to be sufficient to explain hunting in West African chimpanzees (at least at Táï: Stanford 1998) but only if the carcass is divided among the hunters. Given that 47% of individuals sharing in a carcass appear to cheat the system (Boesch and Boesch-Achermann 2000), the possibility that other factors influence hunting needs investigation. The “hunting-as-risk-assessment” hypothesis might apply to the monkey-hunting specialists of the Táï Forest, as is appears to apply to the Ngogo chimpanzees, although it will be necessary to disentangle “hunting-as-risk-assessment” from “meat-for-allies” in testing the relative importance of these two ideas at both sites.

14.7 Conclusions

Chimpanzees are not the only primates that hunt vertebrate prey. Baboons (Papio spp.) also hunt opportunistically, targeting small ungulates (Morris and Goodall 1977; Strum 1987). Among New World primates, capuchin monkeys (Cebus spp.) prey upon a variety of species with Cebus capucinus, perhaps best studied, showing a focus on squirrels, infant coatis, and birds (Rose 1997; Rose et al. 2003), while some squirrel monkeys (Saimiri sp.) hunt bats (Boinski and Timm 1985; Souza et al. 1997).

Among the great apes, vertebrate predation appears to be rare or absent among gorillas and orangutans, but bonobos (Pan paniscus), the phylogenetic sister species to chimpanzees, do hunt vertebrates. Recorded prey species including black and white colobus (Colobus angolensis), red-tailed monkeys (Cercopithecus ascanius) (Sabater Pi et al. 1993), flying squirrels (Kano and Mulavwa 1984), and forest duiker (Cephalophus spp.) (Hohmann and Fruth 1993; Fruth and Hohmann 2002). Hunting by bonobos typically occurs at a lower rate than in chimpanzees: Fruth and Hohmann (2002) report only 9 kills in 46 months of observation, 7 of which were duiker.
While hunting is thus not unique to chimpanzees among the primates, it does appear to be a ubiquitous aspect of their behavior, occurring in all populations studied thus far. The picture that has emerged from these studies is one of diversity but with some common themes. Across populations, hunting is a predominately male activity. Chimpanzees hunt a variety of vertebrate prey, but there is a common focus on medium-sized mammals, particularly primates, and especially colobus monkeys. Red colobus appear to be the preferred prey, although the species (and body size) of red colobus varies across Africa. Chimpanzees appear to impose significant predation pressure on their main prey species, but the intensity and frequency of hunting varies between populations and from month to month within single communities. Hunting is typically opportunistic on encountering the prey, although there is some evidence of chimpanzees searching for prey. Hunts can be solo or group efforts, and the degree to which individual chimpanzees hunt together varies between East and West African populations. This appears to be related to the way the kill is divided following the hunt. In West Africa, the kill tends to be shared according to participation in the hunt and individual hunters collaborate, taking different roles, whereas in East Africa, the kill is typically divided tactically, with the male in possession of the carcass trading parts of it with females for sex, providing the meat after copulating, or with other males for future coalitional support and group hunts are more akin to individual efforts to secure prey occurring simultaneously.

It is important to recognize that this picture comes from detailed systematic studies of only a handful of communities (Gombe, Mahale, Tai, Ngogo). Comparable systematic studies of hunting by chimpanzees in other populations are lacking, although some data are available from almost every population studied. In addition, much of the research effort has focused on chimpanzees and red colobus monkeys. Far less is known about chimpanzee hunting of other species and the nature and importance of hunting in populations that are not sympatric with red colobus. Certainly, chimpanzees without red colobus to hunt appear to hunt less frequently (Basabose and Yamagiwa 1997; Newton-Fisher et al. 2002), and it is unclear what impact low levels of hunting, providing fewer carcasses to share and consume, have on patterns of chimpanzee behavior. If one goal of studying chimpanzee hunting behavior is to shed light on the behavioral ecology of the hominines then addressing the shortcomings in current data is a priority.

The ubiquitous nature of chimpanzee hunting, the common occurrence of food sharing, and the diversity in the patterns of these behaviors, together with the close phylogenetic relationship between chimpanzees and humans, ensure that consideration of chimpanzee hunting is essential in any discussion of the role played by meat eating and food sharing in the behavioral ecology of early
hominine species. The radiation of early hominines encompassed a number of species with different morphologies, and it seems likely that these hominines showed both within and between species variation in habitat and behavioral ecology (Foley 1997). The chimpanzee-red colobus system may be a useful model for some of this variation, but it remains necessary to understand the role of hunting and meat eating across chimpanzee populations, including those with an impoverished resource base. Already it is clear that different populations target different arrays of species, specialize or generalize their choice of prey, and hunt and use meat in different ways. Future studies of new populations are likely to increase this picture of diversity, and systematic tests of the hypotheses for hunting and meat sharing will clarify both why chimpanzees hunt and the importance of this behavior for the study of human evolution.

References

Busse CD (1977) Chimpanzee predation as a possible factor in the evolution of red colobus monkey social organization. Evolution 31: 907–911


Kawabe M (1966) One observed case of hunting behavior among wild chimpanzees living in the savanna woodland of western Tanzania. Primates 7: 393–396


Kortlandt A (1972) New perspectives on ape and human evolution. Stichting voor Psychobiologie, Amsterdam, p 100


Teleki G (1973) The predatory behaviour of chimpanzees. Bucknell University Press, Lewisburg
15 Cooperation, Coalition, and Alliances

Charlotte K. Hemelrijk · Jutta Steinhauser

Abstract

In primates, cooperative acts have been observed such as communal rearing of offspring, cooperative mobbing of predators, supporting others in fights, and grooming others. Grooming builds up a social bond between the partners, helps in repairing relationships, and produces all kinds of benefits for the groomee, such as the reduction of parasites, and of tension. Although the costs for the groomer are low, it has been regarded as an altruistic act and therefore is expected to be preferably directed toward kin or to be repaid by being reciprocated or exchanged for another service (e.g., support in fights, help in rearing offspring in the case of communal breeding systems, or access to some object, such as food, or some individual such as a female, an infant, or members of another group).

The formation of coalitions may result in the maintenance or the increase of the dominance of an individual, in the expulsion of certain individuals from a group, in taking over a group, in the defense of the home range against other groups, in getting access to estrus females, and in the protection of an infant or adult female. The degree of cognition involved in coalitions is unclear.

Which members of a group cooperate differs from species to species; it may be influenced by genetic relations, by the size and the composition of the group (the sex ratio), the degree of competition, and by the distribution of food.

15.1 Introduction

Cooperation in primates varies greatly among members of a group. For instance, individuals groom the fur of others, they help others in fights, collect food together (for communal hunting in chimpanzees, see Volume 2 Chapter 14), share food, and may help in raising the offspring of others. Furthermore, group members cooperate against danger from the outside. They mob predators together and form coalitions to defend their home range against other groups.
For a long time behavioral acts, such as coalition formation, grooming, and food sharing, have been regarded as “altruistic” (costing the actor more than it receives) and therefore, the main explanations have been the theories of kin selection (Hamilton 1964) and reciprocal altruism (Trivers 1971). But in the course of time, it has become increasingly clear that as a rule these supposedly altruistic acts are beneficial for both cooperation partners. For instance, grooming a high ranking individual implies that the groomer does not suffer attacks during grooming. Furthermore, whereas coalitions were originally supposed to require high cognitive abilities, it is increasingly acknowledged that these patterns may arise from simple behavioral rules (as for cognitive mechanisms, see also Volume 2 Chapter 17). Besides, cooperation depends on the social system and the kind of primate involved (such as Old World monkeys versus New World monkeys). We intend to treat these aspects below.

15.2 Social system

Cooperation among individuals of a group depends on the species and its social system. There are many species of primates, and they live in many different kinds of social systems, as solitary individuals, monogamous pairs, single-male groups, multimale groups, or fission–fusion systems (Chapter 12). In group-living species with many females, the males usually migrate and the females remain in their natal group for life (female philopatry), e.g., baboons, macaques, and vervets. Wrangham (1980) refers to these species as “female-bonded,” because the females are more kin-related than the males. In such groups, female social relationships and cooperation are developed much further than among the males of the group and also further than among females of species that are not female-bonded, the so-called female-transfer species. In chimpanzees, for instance, males stay together and females migrate. Here, relationships among males are more cooperative than among females (male-bonded). Greater cooperation among the resident than the migrating sex has been attributed to the closer relatedness of the resident sex. Evidence for closer relatedness in the resident sex is found among macaques (Ruiter 1998) and chimpanzees (Goldberg and Wrangham 1997). Although it has also been argued that male chimpanzees that cooperate are more often more closely related than those that do not, this is not supported by evidence from DNA-typing methods (Goldberg and Wrangham 1997). Furthermore, social relationships differ between Old World monkeys and New World monkeys: among Old World monkeys they are more developed (Dunbar 1993). The causes of these differences are unknown.
Grooming occurs in all primate species (Goosen 1987), and of all affiliative social acts it is the one that is displayed most frequently. It consists in picking through the fur to remove parasites and to clean small injuries. An individual may clean its own fur (autogrooming) or that of another (allogrooming). Because allogrooming is a social act, it has sometimes been questioned whether it has any cleaning function at all. That allogrooming actually aims at cleaning is shown by Zamma (2002): Japanese macaques groom more often those spots of others and of themselves that tend to house more lice and eggs. Furthermore, in his study of 17 spots on the bodies of 19 species of primates, Barton (1985) has shown that individuals groom others particularly at spots that they themselves cannot easily reach. Therefore, spots on the skin that are groomed more often by others are groomed less often by the individual itself and vice versa.

Since in species that live in larger groups, individuals spend more time grooming, grooming clearly also has a social function (Dunbar 1991). Note that this correlation with group size appears more clearly among Old World monkeys than among New World monkeys. This may arise because coalition formation is more important in Old World monkeys and grooming may be helpful in building up alliances. In line with this, it has been found that individuals groom more often those partners they also support more frequently. This has been observed in several species such as female chimpanzees (Hemelrijk and Ek 1991), male chimpanzees (Watts 2002), gorillas (Watts 1997), baboons (Seyfarth 1976; Smuts 1985b), and male bonnet macaques (Silk 1992b).

Grooming relationships are influenced by the composition of the group, namely the sex ratio, for instance, among males of the genus Macaca (Hill 1994) and among females of a number of female-bonded species (Hemelrijk and Luteijn 1998). Food provisioning led to a female-biased sex ratio in a number of groups of species of the genus Macaca (Hill 1994, 1999). This arose because provisioned food was offered in clumps and thus led to stronger competition, and this drove several males away. Groups that were not provisioned had a more equal sex ratio, because in these groups competition for concentrated food sources was less, which may have allowed males to be friendlier among themselves and groom each other more often. Further, more grooming among males in the case of an even sex ratio may arise because the number of males to be groomed is greater and the number of females to groom with is smaller than in groups with a female-biased sex ratio, and the higher number of potential male partners to affiliate with may result directly in more affiliation among males. Grooming relationships among females seem to be influenced by competition for access to males; in a
large comparative study of female-bonded species of primates, Hemelrijk and Luteijn (1998) discovered that the degree of reciprocation of grooming among females increases with the increase in the relative number of adult males in the group. This was attributed to female competition for access to males; the fewer males present, the stronger the competition among females to affiliate with males, and this competition hindered females from building up good relationships among themselves. This argument is supported by the fact that grooming reciprocation increases more strongly with sex ratio among females in a single-male group (where sex ratio depends on group size) than in a multimale group; even if the number of males increases in multimale groups, this increase is not entirely to the profit of the time the males have available for females. There are two reasons for this: first, because males will intervene in interactions of other males with females and second, because males will interact among themselves, which reduces the time available for positive social interaction with females. In single-male groups, however, interactions and interventions among males are lacking. That the degree of grooming reciprocation may be used as a measure of female-bonding is supported by the fact that grooming reciprocation is higher in species that are female-bonded than those in which females transfer.

The positive effect of grooming on relationships is supposed to be a reduction of tension, increase of trust, and restoration of the relationship after a fight. As regards tension, grooming calms and relaxes the groomee (Terry 1970; Goosen 1975): its heartbeat slows down (Boccia et al. 1989) and the rate at which it shows a displacement behavior, such as scratching, decreases (Schino et al. 1988). Keverne and coauthors (1989) have shown that being groomed is felt to be pleasurable for it increases the concentration of endorphins in the brain. Grooming is supposed to maintain relationships in the light of competition, because hamadryas females with an established relationship are observed to groom each other more often if a dyad is accompanied by others than if the dyad is temporarily separated in cages (Stammbach and Kummer 1982).

Grooming is also used to restore a relationship: often after fights, the frequency of grooming and other affiliative behavioral acts increase compared to the situation when no fights take place. This is known as “reconciliation” and has been shown to occur in species of all major radiations of primates (Aureli and de Waal 2000; Aureli et al. 2002). Being attacked implies that there is a high chance that more aggression will follow. Postconflict reunion reduces this aggression and restores the relationship. This function appears from the elegant experiments by Cords (1992). She determined at what distance pairs of long-tailed macaques could drink next to each other without trouble. Then she showed that, after aggression among the members of a pair, its ability to jointly exploit the
resource was seriously reduced. If, however, after such a conflict, a friendly reunion took place, the use of the resource was completely restored to normal.

Relationships are, however, not always damaged by aggression; the damage depends on the context in which the aggression takes place. In the case of competition over food, the relationship keeps its status quo even without reconciliation. Furthermore, the occurrence of reconciliation depends on the value of the relationship. According to the “valuable relationship hypothesis,” reconciliation particularly occurs in relationships of great value (Aureli and de Waal 2000). This theory is supported by the following experiment by Cords and Thurnheer (1993): when macaque partners are obliged to cooperate with each other to obtain food, they reconcile three times more often than when cooperation is not necessary. In line with this, reconciliation has been shown to occur more often in those relationships that are characterized by a high frequency of support (such as in macaques among members of a matriline; and in gorillas reconciliation occurs in the cooperative relationship between the sexes rather than among females [Watts 1995a, b]). Further, in general, more friendly postconflict reunion occurs among those individuals that exchange high levels of friendly behavior. Note that reconciliation by juveniles is already performed in the same way as that by adults, thus cognitive requirements are slight (Aureli et al. 2002).

15.2.1.1 Grooming: kin selection, reciprocation, and exchange

Grooming does not only lead to social bonding but may also be considered an altruistic trait because of the costs to the actor and the advantage for the receiver. Although its cost (expenditure of energy) is low (Wilkinson 1988), grooming may cost time that might be used for: (a) vigilance and (b) foraging. Two studies report a decrease in vigilance, one among captive rhesus monkeys and the other among wild blue monkeys. In rhesus macaques, mothers become less vigilant during grooming and consequently their infants were more often harassed by group members (Maestripieri 1993). Blue monkeys became significantly less watchful of predators when grooming than when foraging or resting (Cords 1995). Grooming does not diminish time for foraging (Dunbar and Sharman 1984): in two species of baboons (olive baboons and gelada baboons), increased foraging time was associated with a decrease in the length of time spent on resting, but time spent on grooming remained the same. This may be an indication of the importance of grooming. Indeed, baboons and macaques devote up to 20% of their time to grooming others (Dunbar 1988).
Within the framework of grooming as an altruistic act, the distribution of grooming partners can be explained either by the theory of kin preference (Hamilton 1964) or by reciprocal altruism (Trivers 1971). In support of kin selection, the most intense grooming bonds are found between mother and offspring, and in general in most primates, individuals aim their grooming primarily at their kin (Gouzoules and Gouzoules 1987; Schino 2001).

When altruistic acts are directed toward unrelated individuals, the expectation is that something should be received in return (Trivers 1971). During a certain part of their grooming bouts—ranging from 5%–7% for *M. radiata* (Manson et al. 2004) to 74% in *Callithrix jacchus* (Lazaro-Perea et al. 2004)—partners groom each other alternately. Recent models suggest that parceling of grooming bouts in small periods, in which the role of actor and receiver alternates, is a method of achieving reciprocation (Connor 1995). In grooming bouts of female chacma baboons, where both partners groom each other in turn, the total grooming duration by both partners is indeed significantly correlated between bouts (Barrett et al. 1999, 2000). Similar findings were made in white-faced capuchin monkeys and bonnet macaques (Manson et al. 2004) but not in Japanese macaques (Schino et al. 2003). Furthermore, it was argued that the time during which an individual grooms another should increase as a sign of the increasing trust among partners (model of “raise-the-stakes,” Roberts and Sherratt 1998). Increasing bout lengths have not, however, been confirmed in empirical studies of either capuchin monkeys or baboons (white-faced capuchins, Manson et al. 2004) (chacma baboons, Barrett et al. 2000). Instead in chacma baboons, bout length even decreased over time.

Grooming may either be reciprocated for its own sake or interchanged for another service, e.g., support, reduction of aggression, or access to something or someone (such as a female, an infant, or another group) or support in rearing offspring (in communal breeding systems). Here, a major problem is how to define reciprocation operationally. Reciprocation and interchange may be considered as a correlation between the number of times each individual gives something to a partner and how often it receives this service from him/her in return. This summed value over a period of time may be studied at the group level, the so-called actor–receiver model (Hemelrijk 1990a, b). Reciprocation in grooming occurs in many species, for instance, among both males and females in chimpanzees in captivity (Hemelrijk and Ek 1991), and among male chimpanzees under natural conditions (Watts 2000a), among female Samango monkeys (Payne et al. 2003), blue monkeys (Rowell et al. 1991), baboons (Seyfarth 1976), marmosets (Lazaro-Perea et al. 2004), female Japanese macaques (Schino et al. 2003), and gorillas (Watts 1994). Such a correlation of reciprocation may, of course, occur as a side effect of other correlations. For instance, when
higher-ranking individuals groom others more often and when everyone grooms others more often according to the rank of the partner, grooming reciprocation follows automatically (Hemelrijk 1990b). To exclude such alternative explanations, partial matrix correlations are useful (Hemelrijk 1990a). Both in chimpanzee males and females (Hemelrijk and Ek 1991; Watts 2002) and in savannah baboons (Seyfarth 1976; Hemelrijk 1991), grooming reciprocation remained manifest even after partialling out the effect of other variables such as dominance and support. In other studies, grooming reciprocation was present while controlling for kinship (hamadryas baboons, Stammbach 1978) (vervet monkeys, Fairbanks 1980), (Japanese macaques, Muoryama 1991). Only in a few studies was no reciprocation of grooming observed (bonobos, Franz 1999).

Apart from being reciprocated, grooming may also be exchanged for other services. For instance, Seyfarth (1977) argues that higher-ranking females are more attractive to groom because from them more effective support in fights can be expected in return. Since females will compete to groom the highest-ranking partners, and since higher-ranking females will win this competition, each female will in the end groom most frequently with those partners adjacent in rank and be groomed most often by those ranking just below her. Seyfarth used this model to explain the observation that in several female-bonded primate species, such as baboons (Seyfarth 1976), vervets (Fairbanks 1980), and stumptail macaques (Estrada et al. 1977), females aimed at grooming up the hierarchy and mainly at those that were next in hierarchy (Seyfarth 1980). Since then, these patterns have statistically been studied in many species. In a number of them, particularly Old World monkeys (such as certain species of macaques, e.g., rhesus monkeys [Kapsalis and Berman 1996], chimpanzees [Hemelrijk and Ek 1991; Watts 2000b], and bonobos [Franz 1999; Vervaecke et al. 2000]), these patterns were, at least partly, confirmed but in others evidence is lacking, for instance, in female langurs (Borries et al. 1994), blue monkeys (Cords 2000, 2002), and in New World monkeys, such as wedge-capped capuchins, Cebus olivaceus (O’Brien 1993), and tufted capuchins, Cebus apella, in both wild (Di Bitetti 1997) and captive colonies (Parr et al. 1997). There was even a trend against grooming higher-ranking animals because individuals groomed down the hierarchy among capuchins and in callitrichids, (Lazar-Perea et al. 2004). In callitrichids, this is suggested to have a function in the communal breeding system: the breeding female (i.e., the alpha-female) uses grooming to make lower-ranking individuals stay in her group in order to help her bring up her young.

Furthermore, the relation between grooming and the receipt of support is doubted. Although correlations were found in studies of several species, such as vervets (Seyfarth 1980), baboons (Seyfarth 1976; Hemelrijk 1990a), female chimpanzees (Hemelrijk and Ek 1991), male chimpanzees (Watts 2002), bonobos
(Vervaecke et al. 2000), capuchins (O’Brien 1993), and one group of bonnet macaques (Silk 1992a), they were lacking in an earlier study of the same group of bonnet macaques (Silk 1982), in rhesus macaques (de Waal and Luttrell 1986) and female baboons (Silk et al. 2004). The relation is supported by two experimental studies dealing with vervets and long-tailed macaques. Seyfarth and Cheney (1984) recorded a call of vervets that seems to solicit support from others. They played it back to individuals of a natural colony of vervets that had recently been groomed by the caller and to others that had not. The duration during which individuals looked up at the speaker was considered to be an indication of the tendency to support the caller. Among nonrelatives, individuals looked at the speaker longer when the caller had recently groomed them. In this experiment it remains uncertain, however, whether looking up at the speaker actually indicates a readiness to support him or her. Therefore, in an experiment with long-tailed macaques (Hemelrijk 1994), actual support was measured directly. Trios of females were separated from the group. After two high-ranking individuals had been given the opportunity to groom, a fight was provoked between one of them and a low-ranking female. The frequency with which the third high-ranking female intervened in the fight was counted. The third female appeared to support only the other high-ranking female, and she did so more often after she had been groomed by her than if not. This supports the notion of a relationship between being groomed and supporting. It is not definitive evidence for an exchange, however, because being groomed may increase the tendency to support in general, even on behalf of those by whom the supporter was not groomed at all. Furthermore, individuals appear to support the aggressor but not the victim; therefore, it is as yet unknown whether a similar association with grooming holds also for victim support (which is more risky). Besides, these experiments do not show whether varying amounts of grooming lead to varying amounts of support.

Henzi and Barrett (1999) suggest that the receipt of support is not the major benefit of grooming because grooming occurs also among females that do not support at all, for instance in certain groups of chacma baboons (in the Drakensberg). Instead, they argue that the short-term benefit of grooming is the decrease of risk of aggression and harassment from others during the grooming bout itself (as suggested for bonnet macaques [Silk 1982] and capuchin monkeys [O’Brien 1993]), because support is rare in female-bonded species, although those females groom each other. Further, Henzi and Barrett (1999) argue that the degree to which grooming should be reciprocated or exchanged for something else depends on the competitive regime; when resources are widely distributed and cannot be monopolized, competition is weak, individuals equal each other in power, and grooming should be reciprocated. If resources can be monopolized, however, competition is intense, power differences are great, and grooming should be
exchanged for increased access to resources. In line with this, in a comparison between groups and in a study of the changes in the same group over time, grooming appears to be reciprocated if competition is weak rather than intense (Barrett et al. 1999, 2002). An alternative explanation for this is found in the fact that in both cases, intense competition is associated with a sex ratio that is more skewed toward adult females. Weaker grooming reciprocation in groups with more females has been discovered in several primate species by Hemelrijk and Luteijn (1998) and is attributed to stronger competition among females for access to males. Note that differences in degree of competition for access to males may also explain the pattern of grooming reciprocation in the baboons studied by Barrett and coworkers.

Grooming is also supposed to be exchanged for access to food. The best evidence for this comes from an experiment with long-tailed macaques by Stammbach (1988). Stammbach trained individual members of a group to become experts in operating a food apparatus. During the period in which it was the expert, this individual appeared to be groomed significantly more often.

Note that “food sharing” in primates almost exclusively means passive tolerance toward others when others take away a bit of food and that active giving is extremely rare (McGrew 1992). In a food exchange experiment among captive chimpanzees, de Waal (1997) found some evidence that females allow others to take away food more easily if they have been groomed by them in the preceding 2 h than if not. Food sharing and its reciprocation seems, however, largely a matter of mutual tolerance rather than intentional reciprocation, as is shown in experiments with brown capuchin monkeys (Waal 2000).

If males groom females, this is further supposed to increase a male’s access to mating partners, for instance in chimpanzees, baboons, and rhesus monkeys. In chimpanzees, this has been regarded as a kind of “bargaining for sex” (Goodall 1986). However, chimpanzee males groomed those females more often with whom they mated more frequently mainly during the period of female tumescence, and this relation neither resulted in a long-term bond (Hemelrijk et al. 1992) nor led to more offspring from the females that were groomed more often by the male (Hemelrijk et al. 1999; Meier et al. 2000). Hence, male grooming of females may simply function to calm down the male’s aggressive tendency or the tendency of the female partner to flee, and therefore, it need not be considered as a kind of exchange or currency. Furthermore, male rhesus monkeys mainly groom females during the mating season, and captive females prefer males who groom them most (Michael et al. 1978); yet, there are no long-term reciprocal bonds between the sexes (Maestripieri 2000). Long-term sociopositive relationships between males and females have been described, however, for savannah baboons (Smuts 1985b; Seyfarth 1978a, b).
In several species, grooming is used to get access to newborn infants, e.g., in chimpanzees, baboons (Rowell 1968), patas monkeys (Muroyama 1994), and moor macaques (Matsumura 1997). Furthermore, if low-ranking female chacma baboons want access to a newborn from a higher-ranking mother, they need to groom the mother longer in proportion to the size of the difference in rank (Henzi and Barrett 2002).

15.3 Coalition formation

A coalition (and alliance) is a coordinated attack by two or more individuals (the coalition partners or allies) on one or more opponents, the so-called targets (Chapais 1995). Coalitions may start in several ways: two individuals may attack a common victim, a coalition partner may spontaneously participate in an ongoing fight, or it may join after being enlisted by one of the combatants. Here, we deal only with those coalitions that are targeted at other group members and not at other groups (for coalitions against other groups see Section 15.5).

Several types of coalitions are distinguished on the basis of their form and their effect (Chapais 1995; van Schaik et al. 2004). As regards form, the distinction is between coalitions of members: (a) that rank above the target (called “all-down”), (b) that rank below the target (called “all-up”), or (c) of which one ranks above the target and another below it (called “bridging” coalitions or alliances). As regards effects, a distinction is made between: (a) alliances that reinforce the existing rank-order, and therefore are “conservative”; (b) coalitions that cause one individual to change rank and thus are “rank-changing”; or (c) coalitions that cause more individuals to change rank, e.g., when two lower-ranking individuals defeat a top male, and thus are “revolutionary.” Coalitions usually involve three individuals (a triad), but more individuals may participate (a polyad).

As regards the cognition involved in the formation of coalitions, opinions differ. Harcourt (1988) suggests that primates form coalitions for strategic reasons and that they must take into account a complex set of information about their own power and that of their allies in comparison to that of their opponent and allies and so on. Along these lines, in a comparative study between a captive group of long-tailed macaques and one of chimpanzees, coalitions of chimpanzees appeared to be more frequent and larger than those of macaques, and this was considered as an indication of their greater cognition (de Waal and Harcourt 1992). Others argue, however, that coalition behavior may develop with little planning and anticipation of the results because individuals may passively learn to recognize the advantage of joining forces (Chapais 1995). Along similar lines, the pattern of coalitions in sooty mangabeys may result from simple behavioral
rules such as “support the higher-ranking individual in a conflict” and “solicit support from potential allies that outrank yourself and the target” (Range and Noe 2005). Even more complex cognition has been suggested to be involved in reciprocation of support. This will be treated below.

During fights, individuals may display so-called enlisting behavior by which they seem to try to attract others into the fight. This may consist of a rapid movement of the head between the opponent and the individual from whom help is requested, called “headflagging” in baboons and “pointing” in long-tailed macaques (de Waal et al. 1976; Noë 1992). In chimpanzees it is part of “side-directed behavior” (de Waal and van Hooff 1981), which consists of several behavioral actions that are shown by fighting animals to a third individual that is not (yet) involved in the fight. One action is the so-called “hold-out-hand” gesture in which an individual stretches out its hand toward a potential helper (de Waal and van Hooff 1981). The effectiveness of such solicitation behavior in chimpanzees is unclear. In contrast to the positive results by de Waal and van Hooff (1981), in another study of the same colony by Hemelrijk et al. (1991) most support was obtained without preceding side-directed behavior and it was clear that side-directed behavior was not a precondition for acquiring help in fights. Besides, side-directed behavior was rarely followed by support, and there was no indication that it increased the chance of receiving help. Note that also if the analysis was confined to cases of hold-out-hand behavior, hold-out-hand appeared not to result in obtaining support. This is in line with the experimental observations of chimpanzees by Hare and coauthors (2004): they discovered that chimpanzees are unable to understand pointing toward an object (Tomasello et al. 1997) but that chimpanzees easily anticipate the stretching arms of those who want to take away something that is of interest to them. Thus, it appears that they are better equipped to compete than to cooperate. Side-directed behavior is mostly displayed by females when they are threatened and is significantly concentrated on higher-ranking individuals; thus, side-directed behavior may be beneficial to the soliciting individuals in the sense that it tends to bring them nearer to a high-ranking individual, which may have a protective effect as a threat to the original opponents (Hemelrijk et al. 1991).

In interventions in fights, it is usually the aggressor and the winner of the fight that is supported. This is less risky than supporting the victim. Notable exceptions are mothers supporting their offspring (see Section 15.3.1.2) and the “control role” of the alpha-male. For instance, in a captive colony of Japanese macaques (Watanabe 1979), the alpha-male more often than other males supports aggresses, in particular babies and youngsters, against adults. This is called a “control role.” A similar control role in the form of supporting losers is described for the alpha-male in gorilla groups by Watts (1997). Here, males
intervene in fights among females. Because this may promote egalitarianism among females, the male may use it to keep females in his group. Male intervention hinders, however, the formation of alliances among females.

15.3.1 Functions of coalition

As regards their function, Smuts (1987) distinguishes a number of main types of coalitions:

a. Coalitions to take over a single-male group: This has been reported for males of grey langurs (Hrdy 1977).

b. Coalitions among males to get access to an oestrous female that is in consort with another male: These coalitions do not affect dominance relationship and are common in savannah baboons (Smuts 1985a). Coalition partners are typically of middle rank (Noe and Sluijter 1995). Packer (1977) found that the male that enlisted help from another was the one to obtain the female, but in a later study Bercovitch (1988) showed that males that solicited were as likely to obtain the female as those that were solicited, and Noë (1992) himself observed that both partners of a coalition may enlist each other’s help simultaneously.

c. Coalitions to repel outside males: This has been reported among females as well as among males. Coalitions among males occur between groups of species living in single-male groups such as gorillas, hamadryas baboons, and gelada baboons. Coalitions among males of a single group have been described in multimale groups of chimpanzees, red colobus monkeys, spider monkeys, white-faced capuchin monkeys, and sometimes also among savannah baboons. In chimpanzees (Wrangham 1999; Wilson and Wrangham 2003) and white-faced capuchin monkeys (Gros-Louis et al. 2003), male coalitions may even result in killing an adult of another group.

d. Coalitions among females or between a male and a female to protect infants: Although the main protectors of youngsters are their mothers, in all species virtually all group members will defend an infant if it is in danger (for instance, against an attack by an adult male).

e. Coalitions among females to protect an adult female against an attack by a male: Mobbing a male to protect a female may be useful for females (even unrelated ones) since it warns males that hostility to females is risky.

f. Coalitions to increase the dominance position of one or both member(s): Here, we will treat results for youngsters and for adults separately.
15.3.1.1 Adults

Coalitions to increase the dominance position among adults have been reported for Japanese macaques, rhesus, Barbary macaques, stump-tailed macaques, mantled howlers, red howlers, chimpanzees, and gray langurs.

When females are observed to attack males together, they are assumed to increase their dominance over males. This is mentioned for Japanese monkeys, rhesus monkeys, and bonobos (Chapais 1981; Thierry 1990; Parish 1994), and indeed in these species certain females are dominant over certain males. Such coalitions are not a precondition for female dominance, however, because a spatial model of individuals that group and compete (via dominance interactions) shows that female dominance can also arise only from competitive interactions in the absence of coalitions (Hemelrijk et al. 2003). According to this model, the stronger female dominance in bonobos compared to common chimpanzees may be due to their greater cohesion in grouping. Furthermore, stronger female dominance among rhesus macaques as compared to Celebes macaques (Thierry 1990) may arise from their higher intensity of aggression. Both factors, cohesion and intensity of aggression, increase hierarchical differentiation, and this reduces the size of the initial difference in dominance between the sexes (Hemelrijk et al. 2003).

When males support females, they usually benefit in terms of dominance. For instance, in Japanese macaques, an alpha-male supported females of lower-ranking matrilines against the alpha-female with whom he had an unstable dominance relationship. Similarly, in rhesus monkeys and chimpanzees, females supported the alpha-male against other males, and in vervets support by females influenced dominance relations among males (Chapais 1995).

Among the three top-ranking males in chimpanzees, coalitions may induce changes of dominance in several ways. For instance, among wild chimpanzees, a top-ranking alpha-male (A) had an unstable relationship with the stronger beta (B). When the gamma-male (C) supported the alpha against the beta, (C) rose in rank above (B) (Nishida 1983). One-and-a-half years later, however, support by the beta caused (C) and (A) to reverse dominance again (Uehara et al. 1994). Thus, the gamma-male played out the alpha-male against the beta-male and this happened also in captivity (de Waal 1982).

In captivity, at other times, the beta-male and gamma-male were observed to join against the alpha-male (a revolutionary alliance). Thus, both rose in rank above the alpha-male (de Waal 1982). Such competition among males has sometimes fatal consequences (Watts 2004). Revolutionary alliances have further been described for male langurs and male Barbary macaques (Chapais 1995).
However, in none of these species did males form coalitions to obtain oestrous females.

15.3.1.2 Youngsters

In a number of female-bonded species (Wrangham 1980), there is also a complex support system that provides young females with approximately the dominance rank of the mother. It has been called a “matrilineal dominance system” and a “nepotistic hierarchy,” because the support involves kin. The matrilineal dominance system implies that all daughters rank immediately below their mother and that among sisters the youngest sister has the highest rank (so-called youngest ascendancy). This classical form of a nepotistic dominance hierarchy is found in rhesus macaques, Japanese macaques, and long-tailed macaques. It comes about as follows. First, closely related females (i.e., a grandmother, mother, or elder sister) support juvenile females and cause them progressively to outrank all adult females that rank below their mother. This process has been detected by an analysis of data of observational studies, and it has also been experimentally demonstrated by Chapais and coworkers in admirable studies on colonies of Japanese macaques and long-tailed macaques (Chapais 1995, 1996; Chapais et al. 2001; Chapais and Gauthier 2004). They created subgroups of three or six juvenile Japanese macaques (with one or two peers of the same matriline) and then added an adult. In this situation, a low-ranking juvenile female appeared to be able to outrank peers in the presence of closely related females but not in the presence of more distantly related kin, such as aunts, grandaunts, or cousins, due to lack of support from them. After female dominance over a lower-ranking matriline was “assigned,” it was maintained by mutual support among members of the same matriline against those of a lower matriline: from that time onward, the young females joined opportunistically in ongoing conflicts against lower-ranking females (called “the common targeting principle”). Within matrilines, younger sisters are dominant over older ones (“youngest ascendancy” rule). This is caused by the mother’s support of a younger daughter against her older ones. If the mother is absent, dominant unrelated individuals will also support younger sisters against older ones. Thus, the network of alliances of females extends beyond her kin.

However, in some groups the matrilineal dominance system is incomplete (weakly nepotistic) (Chapais 2004). For instance, youngest ascendancy is lacking in some feral groups of Japanese macaques (Hill and Okayasu 1995), provisioned groups of Barbary macaques (Prud’homme and Chapais 1993), and one captive group of Tonkean macaques (Thierry 2000). Furthermore, in baboons, daughters...
may outrank their mother during adulthood (Combes and Altmann 2001). Chapais and Lecomte (1995) give three explanations of weak nepotism: demographic, (phylo)-genetic, and ecological. The demographic explanation comes from a model by Datta and Beauchamp (1991), who simulated the effects of demography on female dominance relations by comparing two populations that differ in their growth rate. One population is growing fast (with 2.8 offspring per female) and the other is declining (with 0.6 offspring per female). Thus, in the declining population, matrilines are smaller and potential allies are fewer than in the increasing population. Since a mother needs support from one of her dominant sisters to remain dominant over her daughter(s) and this ally is more likely to be lacking in a declining population than a growing one, mothers will more often become subordinate to their daughters. Similarly, for a youngster to become dominant over her older sister, she needs another sister and her mother as allies. These are more often alive in a growing than in a declining population, and according to Datta (1992) this explains the consistency of the youngest ascendency rule in provisioned and expanding groups of rhesus macaques and Japanese macaques and the lower occurrence of outranking older sisters in the declining population of baboons.

A genetic explanation is given to explain the absence of the youngest ascendency rule in provisioned colonies of Barbary macaques and in a feral colony of Japanese macaques, where allies were present. In Barbary macaques Prud’homme and Chapais (1993) suggest this difference may be a genetic one because they discovered that unrelated individuals rarely supported a younger sister against her older sister (although they support her against lower-born females). This differs from what is known of rhesus macaques and Japanese macaques.

An ecological cause is suggested for the absence of the youngest ascendency rule in feral colonies of Japanese macaques (Hill and Okayasu 1995). Due to the wide spatial dispersion during foraging, the frequency of aggression was rare, and consequently, support was rare too.

Note that apart from species that are weakly nepotistic, there are also those that are clearly non-nepotistic such as Hanuman langurs (Koenig 2000). Furthermore, in nonfemale-bonded species, such as gorillas and chimpanzees, matrilineal dominance does not exist and young females rank according to their age and power.

How can we explain why the matrilineal system evolved in some female-bonded species but not in others? In a comparative study of egalitarian and despotic species of the genus *Macaca*, Thierry (2000) attributes this to differences in degree of despotism. He argues that the matrilineal dominance system is more complete in despotic species due to the higher frequency of support among kin (so that more power differentiation develops between matrilines). He believes...
that this also implies a lower frequency of acts of support among non-kin and that in egalitarian species support is distributed the opposite way, that is, it is more frequent among non-kin than among kin. Furthermore, he shows that dominance styles are phylogenetically conserved (Thierry et al. 2000). This begs the question of what caused the start of the interspecific differences in degree of nepotism between egalitarian and despotic macaques. According to Chapais (2004), this originates from a difference in the “strength of competition” due to the distribution of food. In the case of clumped food, supporting others will be more advantageous. Thus, a nepotistic system develops in which there is a high frequency of support of both kin and non-kin (in contrast to only kin as suggested by Thierry). When food is dispersed and causes scramble competition, it cannot be monopolized and support becomes less useful (both among kin and among non-kin).

15.4 Support: kin selection, reciprocation, and exchange

Support in fights, or coalition, is often thought to be “altruistic” because of the costs (in the form of risks) to the actor and the benefits to the receiver.

As regards the benefits of receiving support, it increases the likelihood of winning a fight, as larger coalitions beat smaller ones (wedge-capped capuchin monkeys [Robinson 1988], bonnet macaques [Silk 1992a]).

The cost of coalitions is difficult to estimate, but often it is assumed that one partner bears most of the costs while the other reaps the benefits (presenting thus a case of altruism). If so, the theory of kin selection and that of reciprocal altruism are believed to explain these supposedly altruistic acts.

As regards kin selection, in Old World primates, individuals support kin more often than non-kin, e.g., in pigtails, stump-tails, rhesus, Japanese macaques, chacma baboons, yellow baboons, and gorillas. Furthermore, individuals aid kin more often if they are more closely related (pig-tailed macaques, chimpanzees, and rhesus monkeys).

In cases of reciprocal altruism, support is supposed to be reciprocated or exchanged for something else. Reciprocation of support is found in a comparative study of rhesus macaques, stump-tailed macaques, and chimpanzees by de Waal and Luttrell (1988). Since the authors statistically partialled out effects of proximity, kinship, and same-sex combination, they argue that reciprocation indicates that individuals keep mental records of the number of acts received from each individual and that they match the number of acts they give to what they have received from each partner. However, in this study, the effects of dominance ranks
and grooming behavior are ignored and data over five consecutive seasons were lumped together. Therefore, what seems to be proof of keeping mental records may have been simply a side effect.

Reciprocation appears to be a side effect, for instance, in the long-term study of the same colony of chimpanzees analyzed per season by Hemelrijk and Ek (1991). It involves a sex difference since chimpanzee males reciprocated support whereas females did not. Males, however, only reciprocated support if their hierarchy was stable; if it was unstable, males supported those they groomed (Hemelrijk and Ek 1991). Because of this, and since there was insufficient indication that there was any negotiation for support—because individuals did not significantly comply with requests from others (Hemelrijk et al. 1991)—it seemed that males might have joined in one another’s fights opportunistically in order to attack common rivals. This may have been the cause of the reciprocation among males. Since males may benefit directly from such joint attacks, supporting behavior is selfish (Bercovitch 1988; Noe 1990) and there is no need for the participants to keep records.

As in chimpanzee males, reciprocation of support has been reported in one study of male baboons (Packer 1977) but not in another (Bercovitch 1988). This difference possibly is related to a difference in the stability of the hierarchy (Hemelrijk and Ek 1991).

The sex difference in reciprocation of support is in line with detailed earlier studies by de Waal (1978, 1984), in which he found that male coalitions were mainly opportunistic and only corresponded with their social bonds during periods in which the position of the alpha-male was clear (de Waal 1984). Female coalitions, however, were more stable and always coincided with their social bonds (Hemelrijk and Ek 1991). Thus, whereas male coalitions seem to serve status competition, female coalitions are directed toward protection of kin and affiliation partners.

Furthermore, de Waal and Luttrell (1988) studied reciprocity of “revenge.” By revenge, one means attacking someone while supporting another for the reason that the subject has received similar “contra-support” before. In this study, it was found that revenge is reciprocated only among chimpanzees, not among the two monkey species (i.e., macaques). This is interpreted as if the individuals aim their support against some individual because they have been attacked in a similar way: they have suffered support against themselves. De Waal and Luttrell consider this a sign of a greater cognitive capacity in chimpanzees because chimpanzees keep track not only of acts of support but also of revenge. However, there are two objections against this interpretation. First, they lumped together data of five seasons despite changes in dominance ranks and in the stability of the power of the top-male. In a study in the same colony, but in an
analysis of support data per season, Hemelrijk and Ek (1991) found no reciprocation of revenge. Therefore, the apparent reciprocation of revenge may have resulted from the lumping together of data, and in any case it is unlikely that individuals keep track of acts of revenge over a period longer than one season. Second, reciprocation of revenge has also been observed in monkeys, e.g., bonnet macaques (Silk 1992a), and also among related female gorillas (Watts 1997), which pleads against the hypothesis of the need of high cognition for such reciprocation.

15.4.1 Communal rearing and allomothering

Usually the mother takes care of the infant alone. In callitrichids (tamarins and marmosets), however, everyone (both parents and mature offspring) assists in rearing the newborns by carrying them and provisioning them. Mature offspring postpone their departure from their natal territory and delay independent breeding (Rapaport 2001). In a number of female-bonded species, youngsters are protected and helped in fights so that they rank immediately below their mothers (matrilineal dominance system, see preceding section) and in all species youngsters are protected in fights that are dangerous. Furthermore, in many species, unrelated individuals may nurse and carry youngsters (allomothering); they cuddle the infant, embrace it, groom, and protect it (McKenna 1979). Allomothers are usually young, nulliparous females, ranking below the mother; often they are sisters of the infant. In this way, the allomother learns how to handle an infant, which increases the chance of survival of her own offspring (Lancaster 1971). An advantage to the mother seems to be the shortening of the interbirth interval (Fairbanks 1990) and the increase of her reproductive success (Ross and MacLarnon 2000). On the basis of detailed comparative studies of macaques, Thierry (2004) argues that two assumptions suffice to explain interspecific variation in degree of allomothering in *Macaca*: (a) attraction to infants and (b) constraints of social structure (McKenna 1979). First, all females are strongly attracted to all infants. Second, in certain species mothers protect the infants with greater care than in others and therefore, in these species, allomothering is counteracted. These species are the species that belong to the two despotic grades of macaques, whose aggression is intense and among whom power differences are great. This may cause problems for females when they have to retrieve their infants. In contrast to this, in egalitarian species, power differences are small and aggression is mild. Thus, differences in the degree of allomothering result from a kind of social epigenesis.
15.5 Collective defense of home ranges

Species and groups differ in the way in which they use their home range. Depending on this, fights with other groups may aim at the defense of only one food source (e.g., fruit tree) or of a whole territory (Cheney 1987). A number of species have special intergroup calls that are meant to separate the groups spatially (e.g., in mantled howler, capuchin, mangabey, siamang, yellow-headed titi). Most territorial species, however, have intergroup calls that incite the other group to fight them (e.g., dusky titi, gibbon spp., vervet, colobus).

If actual fights between groups occur, in chimpanzees this may lead to killing an adult of the other group (Wilson and Wrangham 2003; Wilson et al. 2004).

Usually, males are more active than females in fights between groups. In macaques, however, females may also participate. In both sexes, higher-ranking individuals participate more often than lower-ranking ones (Cooper 2004).

15.5.1 Behavior against predators

When primates meet a predator, they flee individually (so do large species) or hide (in particular, smaller species). Furthermore, they may protect themselves and their group members in other ways such as: (a) by mobbing the predator, (b) by scanning the environment for early discovery of predators, and (c) by warning other group members.

Mobbing predators has been described for baboons, rhesus monkeys, and all three ape species. Each of these species was observed to attack tigers or lions (Cheney and Wrangham 1987). Scanning the environment has been described for a number of primate species such as red-bellied tamarins (Caine 1984) and chacma baboons (Hall 1960). In a series of experiments, tamarins appeared to scan most frequently during the most dangerous periods of the day and in the presence of the most dangerous stimuli. Further, tamarins appeared to divide the duty of scanning among group members. In relation to protection against predators, the spatial distribution of the individuals of the different sex age categories in progressions has also been studied in baboons (Altmann 1979; Rhine and Westland 1981).

Most primates use alarm calls to warn against predators. Such calls are altruistic in the sense that they are harmful to the sender, because it attracts the attention of the predator to the caller, but beneficial to others that are close by. It has been debated whether kin-selection is the main evolutionary force behind these calls because it is usually kin that is protected by these calls. However, newly immigrant males also tend to call loudly in spite of the fact that they have no
kin-members in the group (Cheney and Seyfarth 1981). Many species of primates use different alarm calls for different predators. Such a differentiation is described for vervets, red colobus, Goeldi marmosets, pygmy marmosets, cotton-top tamarins, and gibbons (Cheney and Wrangham 1987; Dugatkin 1997). Vervets emit different calls when the predator is a leopard, an eagle, or a snake (Struhsaker 1976). In playback experiments of the different calls, vervets appeared to respond to leopard alarms by climbing into the trees; at eagle alarms they looked up, and at snake alarms they looked down. Because of the fine distinction between these alarm calls (almost resembling human language), it has been asked whether alarm calling can be considered intentional warning. Evidence points against this, because vervets continue to give alarm calls after everyone in the group has heard them (Cheney and Seyfarth 1985) and because the intensity of these alarm calls and other protective actions by mothers remained similar whether or not their daughters were informed about the presence of the predator (Cheney and Seyfarth 1990).

References


de Waal FBM, Luttrell LM (1986) The similarity principle underlying social bonding
Hare B, Tomasello M (2004) Chimpanzees are more skilful in competitive than in cooperative cognitive tasks. Anim Behav 68: 571–581
Hemelrijk CK (1990a) A matrix partial correlation test used in investigations of reciprocity and other social interaction patterns at a group level. J Theor Biol 143: 405–420
Hemelrijk CK (1990b) Models of, and tests for, reciprocity, unidirectional and other social interaction patterns at a group level. Anim Behav 39: 1013–1029
Hemelrijk CK (1994) Support for being groomed in long-tailed macaques, Macaca fascicularis. Anim Behav 48: 479–481
Hemelrijk CK, Ek A (1991) Reciprocity and interchange of grooming and ‘support’ in captive chimpanzees. Anim Behav 41: 923–935
Keverne EB, Martensz ND, Tuite B (1989) Beta-endorphin concentrations in cerebrospinal fluid of monkeys are influenced by grooming relationships. Psychendocrinology 14: 155–161
Matsumura S (1997) Mothers in a wild group of moor macaques (Macaca maurus) are more attractive to other group members when holding their infants. Folia Primatol 68: 77–85
Noë R (1990) A veto game played by baboons: a challenge to the use of the prisoner’s dilemma as a paradigm for reciprocity and cooperation. Anim Behav 39: 78–90
Seyfarth RM (1976) Social relationships among adult female baboons. Anim Behav 24: 917–938
Seyfarth RM (1978b) Social relationships among adult male and female baboons. II. Behaviour throughout the female reproductive cycle. Behaviour 64: 227–247


Smuts BB (1985b) Sex and friendship in baboons. Aldine, Hawthorne NY


Watts DP (1995a) Post-conflict social events in wild mountain gorillas (Mammalia, Homi-
noidea). I. Social interactions between opponents. Ethology 100: 139–157
1 Potential Hominoid Ancestors for Hominidae

George D. Koufos

Abstract

The origin of hominids and especially man is a question that has occupied and still occupies scientists. There are a number of discussions and disagreements, and each new finding creates new problems, discussions, and arguments. This article aims to give all available information about the potential hominoid ancestor of the hominids, using all recent data. Certainly, it does not solve the problem but provides an approach summarizing the known data and ideas. Oligocene Aegyptopithecus seems to be the possible ancestor of the Miocene hominoids as it shares some derived features with the early hominoids. The early Miocene hominoids Proconsul is the most possible link connecting the hominoids to the modern great apes and humans. Among the Late Miocene hominoids, Ouranopithecus seems to be closer to the hominids and extant great apes. However, the African Sahelanthropus and Orrorin seem to have some relationship to the Plio-Pleistocene hominids.

1.1 Introduction

The presence, the origin, and the future of man on earth are problems that have long preoccupied humans. Questions like “who is my ancestor,” “what was it like,” or “where do I come from” were among the first for which answers were sought. The need of humanity to find answers to these questions led first to origin myths and later, after the development of science, to theories of human origins. Anthropology and biology give answers based on the comparative anatomy, behavior, and genetics of humans and other animals. The evolutionary approach to understanding human origins seeks to find all the characters that connect humans with apes (our closest relatives) and with the other animals.

Paleoanthropology is the branch of paleontology that addresses the above questions based on the primate fossil record. Paleoanthropologists try to find links connecting various fossil and extant taxa in order to complete the “chain” leading to man. It is difficult and time-consuming work as fossils are relatively
rare and widely dispersed, and in most cases the remains are fragmentary, providing limited data. The discovery of new fossils over the last 30 years and the development of new methods for studying, comparing, and dating fossils, as well as reconstructing their locomotion and paleoenvironment have significantly increased our knowledge about the morphology and evolutionary relationships of primates. One of the main goals of paleoanthropology is to find the ancestor of humans (usually referred to as hominids) from among the common stock of hominids and apes (known as hominoids). The hominoids constitute a group of African and Eurasian forms that lived during the Miocene epoch and are considered to include the common ancestor of humans and apes.

The first known hominoids are from Africa. The earliest possible representative of the hominoids is the latest Oligocene *Kamoyapithecus* from Lothidok (Kenya) which is dated from 27.8 to 23.9 Ma (Leakey et al. 1995). Hominoids are more securely represented by the large sample of Early Miocene proconsulids. The majority of the Early Miocene hominoids is restricted to Africa and is only found in East Africa and the Arabian Peninsula. A large number of taxa are recognized (*Proconsul*, *Afropithecus*, *Heliopithecus*, *Nyanzapithecus*, *Mabokopithecus*, *Rangwapithecus*, *Turkanipithecus*, *Dendropithecus*, *Micropithecus*, *Simiolus*, *Morotopithecus*, *Limnopithecus*, *Kalepithecus*), but most of them are known from few fragmentary fossils. Their size varies from small monkey sized (~3 kg) to large great ape sized (~80 kg), while their locomotion was quadrupedal and arboreal. The majority were fruit or leaf eaters. *Kenyapithecus* appeared at the end of Middle Miocene (~14 Ma) and was probably a hard-object feeder. The uppermost Middle Miocene and Late Miocene hominoids of Africa are rare and poorly known compared to Early/Middle Miocene forms (*Kenyapithecus*, *Samburupithecus*, *Otavipithecus*, *Orrorin*, *Equatorius*, *Sahelanthropus*).

During the end of Middle Miocene, the “story” of hominoid evolution moved to Eurasia, where they were abundant until Late Miocene (~9 Ma). Hominoids migrated to Eurasia and are represented by the widespread taxon *Griphopithecus*. After their arrival, they strongly diversified and several new taxa (*Dryopithecus*, *Pierolapithecus*, *Oreopithecus*, *Ouranopithecus*, *Sivapithecus*, *Ankarapithecus*, *Lufengpithecus*) appeared between 9.0 and 13.0 Ma. The size of these taxa varies from medium to large, while their diet ranges from soft-object feeders, like *Dryopithecus* and *Oreopithecus*, to hard-object feeders like *Ouranopithecus*, *Lufengpithecus*, and *Gigantopithecus*. Postcranial remains are relatively rare and indicate quadrupedal locomotion; however, one taxon is interpreted by some as an early orthograde (Moyà-Solà and Köhler 1996).

Considering that the occurrence of the oldest australopithecines is dated to ~4.5 Ma and the last Miocene hominoids date to the end of Vallesian (~8.7 Ma), there is a large gap of ~4 Ma between them. Until the beginning of the twenty-first
century, no fossils were known from this time interval. The discovery of the African hominoid *Orrorin*, dated to 6 Ma (Senut et al. 2001), is the first evidence to fill this gap. One year later, the skull of *Sahelanthropus*, found in Tchad and dated to 6–7 Ma, offered more data about this unknown time interval (Brunet et al. 2002). More recently, a maxilla of a hominoid from the Tuolian of Turkey (Begun et al. 2003) seems to add one more link in the chain. The complete study of these specimens and their comparison with the previously known material will provide new data relevant to our understanding of the ancestor of the hominids.

The present article aims to give all available data about the potential hominoid ancestor of the hominids. Certainly, it is not the solution to the problem but an approach summarizing the known data and ideas.

### 1.2 Origin of the Miocene hominoids

The origin of the Miocene hominoids is to be found among the Oligocene primates. The division of the anthropoids into two main groups, Platyrhini (New World monkeys) and Catarrhini (Old World monkeys), is thought to have occurred at the Eocene/Oligocene boundary. One of the oldest and best-known catarrhine is *Aegyptopithecus* (*Figure 1.1*), from the Fayum of Egypt with an estimated age from 32 to 35 Ma (Tattersall et al. 1988). Considering that the Eocene/Oligocene boundary is estimated at ~35 Ma, the splitting of the catarrhines and platyrhines coincides with the beginning of Oligocene and bridges the gap between the Eocene Prosimii and the Miocene hominoids.

*Aegyptopithecus* preserves some characters of the Platyrhini such as the absence of the auditory tube in the external ear, but it has two premolars in

*Figure 1.1*
*Aegyptopithecus zeuxis*, skull found in 1966. (a) Lateral and (b) frontal view. Drawn from the photo of Szalay and Delson (1979; fig. 222)
each half of the jaw, a feature that serves to classify it within the Catarrhini, being among the most primitive of the Catarrhini. The main morphological characters of \textit{Aegyptopithecus} (\textit{Figure 1.1}) are an elongated skull with strong prognathism, strong postorbital constriction, large and completely enclosed orbits, weak supraorbital torus, large interorbital distance, a posteriorly developed sagittal crest, well-developed sexual dimorphism in the canines, broad molars with rounded and low cusps, well-developed cingulum, and thin enamel. The braincase is small and its capacity is estimated at \(\sim 30 \text{ cm}^3\) (Conroy 1990). The endocasts indicate a morphology between the primitive and more derived anthropoids (Radinsky 1973). The humerus is relatively stout with a large medial epicondyle and relatively wide trochlea. The morphology of the metapodials and phalanges suggests powerful grasping. The overall morphology of \textit{Aegyptopithecus} indicates quadrupedal climbing, while the morphology of the skull and teeth suggests a “soft-object” feeder. This type of diet fits well with its postcranial morphology indicating a good climber.

The presence of some derived characters in \textit{Aegyptopithecus} has been interpreted in different ways. Tattersall et al. (1988) considered that \textit{Aegyptopithecus} represents an early catarrhine close to the split between platyrrhines and catarrrhines. Simons (1965), focusing on features shared between \textit{Aegyptopithecus} and \textit{Proconsul}, suggested a close phyletic relationship between the two genera. Later, Andrews (1978) reported close similarities between the upper molars of \textit{Aegyptopithecus} \textit{zeuxis} and \textit{Proconsul africanus}, but he could not identify clearly shared derived features. The cranial proportions and the external ear with a clear auditory tube in \textit{P. africanus} are more derived features than those of \textit{Aegyptopithecus} (Szalay and Delson 1979). The primitive robust postcrania of \textit{Aegyptopithecus} with less joint mobility suggest more leaping locomotion, while in the Miocene hominoids, leaping represents a small part of their locomotion (Gebo 1993). \textit{Aegyptopithecus} is a mosaic of primitive and derived characters, linking Eocene anthropoids to modern apes and monkeys and could represent or be close to their common ancestor. However, the available data do not fully support this view. Even its specific name \textit{zeuxis} reflects this fact. It is a Greek word, meaning the way (the “bridge”), which joins or connects two parts. Despite the possible link between \textit{Aegyptopithecus} and hominoids, there is no clear bridge, but in fact a 10.0 Ma gap in the fossil record between \textit{Aegyptopithecus} and the first Miocene hominoids. On the other hand, during Oligocene, several Afro-Arabian taxa of early catarrhines are known (\textit{Catopithecus, Oligopithecus, Moeripithecus, Proplopithecus}), which could be also related to the Miocene hominoids. Among them \textit{Aegyptopithecus} is the youngest and closest to the Miocene hominoids.
1.3 The Miocene hominoids

While the Eocene is the period of Prosimii, and Oligocene that of early anthropoids, the Miocene represents the epoch of the hominoids. Many genera and species of hominoids have been recognized during Miocene. This taxonomic inflation is due either to the limited material or to different opinions between the various scientists, as well as to the tradition among paleoanthropologists to erect new taxa. The hominoids expanded across the Old World from Spain to China and from South Africa to northern Europe (Figure 1.2). During the last 30 years, extensive field campaigns carried out in the Old World by several scientific groups have brought to light numerous fossils of hominoids. In addition, the revision of old material and the new “discoveries” in museum’s collections have increased the Miocene hominoid fossil record. The great increase of knowledge accumulated during the last decades has led to the recognition of several taxa, better defined through the application of cladistic methods and to the recognition that the Miocene was an epoch of high diversification for the hominoids.

During Miocene, several geotectonic movements took place in the Old World causing geographic and environmental changes that imposed either positive (diversification, expansion, migration) or negative (reduction or even extinction) effects on the hominoids. Many new hominoids appeared in Eurasia after their arrival from Africa at the upper part of Middle Miocene. However, at the beginning of Turolian (~8.7 Ma) all these forms disappeared because of increasing dryness during Turolian (Koufos 2006). The new drier conditions were not favorable to the hominoids and probably caused their extinction, with the exception of Gigantopithecus. The latter is also known from the Pleistocene of southern China and Vietnam (Kelley 2002). On the other hand, all these changes (paleogeographic, paleoclimatic) during Miocene were strong factors contributing to the rapid evolution and diversification of the hominoids, leading to more derived and better adapted forms. The increasing severity of ecologic conditions caused the extinction of most Miocene hominoids, but for some of them it was an opportunity to develop some more derived and evolved characters, which allowed their adaptation to the new conditions.

As mentioned above, Aegyptopithecus could be the ancestor of the Miocene hominoids, while the possible common ancestor of humans and apes must be among the stem Miocene hominoids. From this point of view, the hominoids are a very important and interesting group of primates. During the last 50 years, several hominoids have been presented as possible ancestors of the hominids, but the resolution on this problem is still in progress. The problem is complicated and
cannot be solved easily, as the relationships among the fossil hominoids, as well as between the extinct and extant ones, are not fully known. The data for the fossil hominoids in most cases are limited and do not allow clear definitions and comparisons. The number of specimens is another problem. No fossil hominoid is known well enough for the complete knowledge of its morphology and relationships. There are several important time gaps between the various forms, which prevent a clear establishment of their relationships. Finally, the subjectivity of each scientist raises some additional problems. In an effort to evaluate the evidence for
the most probable ancestors of the hominids, the Miocene hominoids, most often considered in this regard, will be discussed in the following.

Knowing the age of fossil hominoids is important for establishing their evolutionary relations. The stratigraphic distribution of the main Miocene hominoids of Africa and Eurasia is given in Figure 1.3. The European land mammal stages and zones are also given in Figure 1.3, as they are mentioned in the text. It is quite clear from Figure 1.3 that Miocene hominoids are well known from the Early–Middle Miocene of Africa, while in Eurasia, hominoids are more common during Late Miocene.

1.3.1 Early Miocene

The Early Miocene hominoids are only known from Africa and the best known taxon is *Proconsul* (Figure 1.4) (Walker 1997). The oldest remains of *Proconsul* (*P. africanus*, *P. major*) were found in Kenya and Uganda and are dated to 19.0–20.0 Ma. Two other species (*P. nyanzae* and *P. heseloni*) are known from Kenya and dated to 17.0–18.5 Ma (Harrison 2002).

*Proconsul* has an estimated body size that varies from that of a small monkey to that of a female gorilla. There is sexual dimorphism in the body size as well as in the size and morphology of the canines. This sexual dimorphism is a primitive feature for *Proconsul*. The nasomaxillary region of *Proconsul* has a primitive internal structure with a large fossa incisiva. In addition, the maxillary processus palatinus is clearly lower than the premaxilla (Ward and Kimbel 1983; Ward and Pilbeam 1983). This nasomaxillary morphology (“African type” according to the latter authors) is similar to that of *Australopithecus, Dryopithecus, Ouranopithecus*, and the recent *Gorilla*. The skull has a moderately short and broad face compared to that of *Aegyptopithecus*, and a broad and relatively rhomboid-shaped nasal aperture, trapezoid orbits with rounded corners, slight supraorbital torus with slightly swollen glabella, and a clear auditory tube like the recent cercopithecoids and hominoids. The cranial capacity is larger than *Aegyptopithecus* and similar to gibbons and recent cercopithecoids, while the endocranial casts indicate similarities to recent gibbons (Szalay and Delson 1979; Walker 1997; Harrison 2002).

The dentition of *Proconsul* preserves primitive features, such as relatively long and broad canines, well-developed cingulum, asymmetric upper premolars with large paracones (buccal cusps), and the presence of a honing facet on the P3. Although the dental morphology of *Proconsul* is well known, there are no clear synapomorphies with either the hominoids or cercopithecoids. In many ways, it is similar to the dentition of Oligocene Fayum catarrhines.
Figure 1.3
Stratigraphic distribution of the main Miocene hominoids

<table>
<thead>
<tr>
<th>Time (Ma)</th>
<th>Chron</th>
<th>Polarity</th>
<th>European Land Mammal Stages</th>
<th>MN-Zones</th>
<th>Epoch</th>
<th>Africa</th>
<th>Europe</th>
<th>Asia</th>
</tr>
</thead>
<tbody>
<tr>
<td>25.5</td>
<td>25.5</td>
<td>E</td>
<td>Oligocene</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23.0</td>
<td>23.0</td>
<td>E</td>
<td>Oligocene</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21.5</td>
<td>21.5</td>
<td>E</td>
<td>Oligocene</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20.0</td>
<td>20.0</td>
<td>E</td>
<td>Oligocene</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18.5</td>
<td>18.5</td>
<td>E</td>
<td>Oligocene</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17.0</td>
<td>17.0</td>
<td>E</td>
<td>Oligocene</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15.5</td>
<td>15.5</td>
<td>E</td>
<td>Oligocene</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14.0</td>
<td>14.0</td>
<td>E</td>
<td>Oligocene</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.5</td>
<td>12.5</td>
<td>E</td>
<td>Oligocene</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.0</td>
<td>11.0</td>
<td>E</td>
<td>Oligocene</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.5</td>
<td>9.5</td>
<td>E</td>
<td>Oligocene</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.0</td>
<td>8.0</td>
<td>E</td>
<td>Oligocene</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.5</td>
<td>6.5</td>
<td>E</td>
<td>Oligocene</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.0</td>
<td>5.0</td>
<td>E</td>
<td>Oligocene</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.5</td>
<td>3.5</td>
<td>E</td>
<td>Oligocene</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.0</td>
<td>2.0</td>
<td>E</td>
<td>Oligocene</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.5</td>
<td>1.5</td>
<td>E</td>
<td>Oligocene</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>1.0</td>
<td>E</td>
<td>Oligocene</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.5</td>
<td>E</td>
<td>Oligocene</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Potential hominoid ancestors for Hominidae
The known postcrania of *Proconsul* are relatively abundant and provide a large set of characters. However, the systematic interpretation of these characters leads to contradictory results. *Proconsul* has anterior and posterior limbs of more or less similar length indicating quadrupedal arboreal locomotion. On the other hand, recent hominoids have a more erect posture (knuckle walking and brachiation). The robusticity of the long bones of *Proconsul* suggests similarities to cercopithecoids, chimpanzees, or to ceboids with less elongated limbs. The distal extremity of the *Proconsul* humerus resembles that of small apes, while the distal extremity of ulna is relatively straight, as in most anthropoids. The carpals indicate a climbing primate, while the phalangeal proportions are very close to macaques, suggesting a quadrupedal form (Walker and Pickford 1983; Ruff et al. 1989; Rose 1992, 1993).

Most authors consider *Proconsul* to be a stem hominoid, close to the origin of the Hominoidea (Hopwood 1933; Pilbeam 1969; Rose 1983, 1992, 1993; Andrews 1985; Begun et al. 1997; Kelley 1997; Ward 1997). All these authors consider that *Proconsul* shares derived characters with the extant hominoids. A different opinion suggests that *Proconsul* is a sister taxon of the recent great apes and humans (Walker and Teaford 1989; Rae 1997; Walker 1997). Neither opinion is accepted by Harrison (2002), whose detailed and critical analysis of key synapomorphies of the face, ear region, and postcrania shared by extant hominoids are interpreted to be lacking in *Proconsul*. Walker (1997) refers to shared derived characters between *Proconsul* and more advanced hominoids, but he also recognizes possible features in the postcrania indicating a primitive catarrhine.
The frontal sinus in the skull of *Proconsul* could be a derived character, shared with the extant hominoids and great apes (Andrews 1992). Although the frontal sinus is absent in Old World monkeys, it is clearly present in New World monkeys, making the interpretation of this character complicated. The absence of a tail is a derived character of the hominoids, and the lack of tail in *Proconsul* (Ward et al. 1991) suggests a relationship to the hominoids. The presence of six lumbar vertebrae rather than seven in *Proconsul* (Ward et al. 1993) indicates a similarity to the hominoids and distinguishes it from monkeys. Several characters of the phalanges could be derived and shared with extant apes (Begun et al. 1994). The morphology of the trapezium and first metacarpal of *Proconsul* could be considered as derived, with several characters shared between *Proconsul* and extant hominoids (Rose 1992). The increased deviation of the ulna of *Proconsul* is a derived character of the hominoids (Beard et al. 1986). The deviation of the ulna is not strong in nonhominoid catarrhines. Rae (1997) studied a set of facial characters in Early Miocene and modern hominoids and found that Early Miocene hominoids, including *Proconsul*, are linked with the modern hominoids and particularly with great apes. *Proconsul* was recognized by Walker (1997) as a morphologically primitive hominoid, sharing several derived characters with the extant hominoids, and having few characters linking it with the extant great apes and humans.

*Proconsul* has a set of characters that link it with the Miocene hominoids or even with the extant great apes and humans but controversies remain. Future research could provide more data for more certain and clear results about its phylogenetic relationships. In addition to *Proconsul*, several other taxa are known from the Early Miocene of Africa that could be related to the extant great apes and humans, and it is quite interesting to give some information about them and their relationships.

The genus *Afropithecus* was recognized in Kenya from localities dated to early Miocene and their age is estimated between 17.0 and 18.0 Ma (Leakey and Walker 1997). *Afropithecus* is characterized by its relatively large size (similar to the recent chimpanzee), long and wide skull, narrow palate, large diastema, narrow and protruding premaxilla, large interorbital distance, asymmetrical orbits, slender supraorbital torus, frontal sinus extending to glabella, lack of the superior transverse torus in the mandible, and an oblique angle between ascending ramus and mandibular corpus. The upper canine has a round basal section and deep mesial groove, while the lower canine is low crowned and laterally compressed. The upper premolars are wide and without lingual cingulum, while the cusps are moderately heteromorphic. The upper molars are narrow with moderately developed lingual cingulum and very thick enamel. The few known postcrania resemble those of
Afropithecus is a large hominoid, whose facial shape resembles that of Aegyptopithecus zeuxis (Leakey et al. 1991), but its postcrania are more derived than Aegyptopithecus and closer to Proconsul (Harrison 2002). The teeth of Afropithecus are more derived than those of the Oligocene forms (enlarge and procumbent incisors, thick canines). These characters as well as the very thickly enameled teeth suggest a “hard-object” eater and thus it is closer to Kenyapithecus. The lack of the latter characters in the dentition of Proconsul indicates a less derived feature, but its derived facial shape is closer to the later hominoids.

Morotopithecus is known from Uganda from a single species, M. bishopi, which was originally described as Proconsul major. Its age is debatable, as the correlation of the fauna and older K⁴⁰/Ar³⁹ datings suggest middle Miocene age, ~15.0 Ma (Bishop et al. 1969; Pickford et al. 1999), while ⁴⁰Ar/³⁹Ar datings suggest an Early Miocene age, older than 20.6 Ma (Gebo et al. 1997). It is relatively small sized, smaller than P. major. The skull is characterized by a long, high, and narrow face, short premaxilla, wide palate, narrow interorbital distance, large diastema, and primitive nasomaxillary region with a large fossa incisiva. There is strong sexual dimorphism in the upper canine. The upper premolars are broad and relatively large. The molars have bunodont cusps, wrinkled enamel, and well-developed lingual cingulum similar to those of P. major (Pilbeam 1969; Andrews 1978; Harrison 2002). The cranial and dental characters of Morotopithecus suggest that it is close to the proconsulids.

The postcrania of Morotopithecus are few but provide some significant characters for the genus. The anatomy of the lumbar vertebrae shares some derived features (robust pedicles, reduced ventral keeling) with the extant hominoids. The scapula (some authors do not agree that it belongs to this genus) has a rounded and expanded upward glenoid articular surface, as in hominoids. Finally, the morphology of the femur and phalanges resemble that of Proconsul and indicates an arboreal form (Walker and Rose 1968; Gebo et al. 1997; MacLatchy et al. 2000; Harrison 2002). On the basis of the postcranial morphology, Morotopithecus could be considered as belonging to Hominoidea (Gebo et al. 1997; Harrison 2002). On the other hand, Andrews (1992) reported that the morphology of the skull and teeth of Morotopithecus are less derived than proconsulids, while the upper teeth share derived features with afropithecines. Morotopithecus seems to belong to Hominoidea from its postcrania morphology, but its cranial and dental characters indicate a proconsulid. It is recognized as a stem hominoid (Gebo et al. 1997; MacLatchy et al. 2000; Harrison 2002);
however, the Harrison (2002) puts the question that the derived features of the vertebral column might be an adaptation to an increased orthogradry.

Several other Early Miocene genera are also included in the family Proconsulidae, but most of them are known from one specimen (maxilla, mandible, or fragments) or isolated teeth. Their comparison with the other extinct and extant hominoids is limited and does not provide clear relationships with them. *Heliopithecus* is known from a maxilla and some isolated teeth from the Early–Middle Miocene of Saudi Arabia (Andrews et al. 1978; Andrews and Martin 1987). The dental characters of *Heliopithecus* suggest it may be synonymous with *Afropithecus*, as well as with *Morotopithecus* (Andrews et al. 1978; Andrews and Martin 1987). On the other hand, Harrison (2002) considers that both *Afropithecus* and *Morotopithecus* are generically distinct and *Heliopithecus* is more primitive than both of them. *Nyanzapithecus* is better represented by several maxillary and mandibular remains from Kenya dated to 13.0–18.5 Ma (Early–Middle Miocene). *Rangwapithecus* is known from the Middle Miocene of Kenya (19.0–20.0 Ma), and *Mabokopithecus* and *Turkanapithecus* from the Middle Miocene of Kenya dated to 15.0–16.0 Ma and 16.6–17.7 Ma, respectively. *Nyanzapithecus* and *Mabokopithecus* have a close relationship and could be congeneric, while *Rangwapithecus* is a sister taxon of *Nyanzapithecus* (Andrews 1978; Harrison 1986a, b, 2002). The facial and postcranial characters of *Turkanapithecus*, shared by *Turkanapithecus* and *Proconsul*, indicate a close relationship between them (Harrison 2002). As indicated earlier, the relationships of all these genera to each other, as well as to the other extinct and extant hominoids, are very uncertain as the available material is poor (sometimes only isolated teeth), thus making comparisons quite doubtful. Despite all these difficulties, doubts, and uncertainties, the ancestor of the Early/Middle Miocene hominoids is included in this group. This is the reason why among them some taxa resemble later hominoids in various derived characters but also retain some primitive features. Derived characters of the face, the teeth, or postcrania of various taxa allow for any one of them to have a close relationship to the younger forms. The group of the Early Miocene hominoids includes some taxa, like *Proconsul*, which seem however to have more similarities to the later hominoids and could be their possible ancestor.

### 1.3.2 Middle Miocene

Compared to the Early Miocene African hominoids, the Middle Miocene ones are few. One of the most important taxa is *Kenyapithecus*, known from Kenya (Leakey 1962) and dated to 14.0 Ma (Kelley and Pilbeam 1986). It is a medium-sized hominoid with clear canine fossa, highly zygomatic arches, clear sexual
dimorphism, strong inferior torus, robust mandibular corpus, less asymmetric P₄, large upper molars without lingual cingulum, and P₃ with a honing facet and buccal cingulum (Ward and Duren 2002). Few postcrania of *Kenyapithecus* are known but include some evolved characters such as a posteriorly directed medial epicondyle and a wide trochlea in relation to the breadth of the capitulum of the humerus (Andrews and Walker 1976). The phylogenetic relationships of *Kenyapithecus* are still debated. *Kenyapithecus* has derived mandibular and dental characters (larger inferior than superior transverse torus, relatively short and wide corpus) shared with extant hominoids. Compared to Early Miocene hominoids, it is characterized by a reduction or absence of cingulum in the molars, an increase of the enamel thickness, large upper premolars relative to the molars, and the more molarized dp₃ close to that of *Gorilla* and *Pan* (McCrossin and Benefit 1997). All these features are shared with *Ouranopithecus*, *Sivapithecus*, and *Griphopithecus*.

*Equatorius* is another Middle Miocene African hominoid found in Kenya, which is dated between 15.5 and 14.0 Ma (Feibel et al. 1989; Ward et al. 1999). It is medium sized with strong sexual dimorphism, more symmetrical upper premolars, reduced or absent lingual cingulum, less heteromorphic lower premolars, thick enamel, and p₃ with honing facet, and variably developed buccal cingulum. It is considered as a primitive hominoid, belonging to a derived clade of *Afropithecus* (Ward and Duren 2002). The morphological characters of the teeth and mandible of *Equatorius* and *Kenyapithecus wickeri* suggest that there is no generic difference between the two taxa, which can be synonymized (Benefit and McCrossin 2000). Begun (2000) questioned the generic distinction between *Equatorius* and *Griphopithecus* from Europe, suggesting evidence of a biogeographic link between the two. Ward et al. (1999) and Kelley et al. (2000) reported that the incisor, canine, and maxillary morphology of *K. wickeri* are more derived than *Equatorius* and distinguish the two genera. The robust mandible and the large inferior torus of *Griphopithecus alpani* (the most common hominoid in Paşları, Tuzkey) are similar to those of *Kenyapithecus*, but the upper jaw is less robust than that of *Equatorius*, while the dental morphology is similar (Andrews and Harrison in press). On the other hand, *K. wickeri* exhibits similarities to the rarely represented hominoid of the Paşalar sample, which are enough to attribute both to the same genus (Martin and Andrews 1993; Ward et al. 1999; Kelley et al. 2000). If all these observations are true then this may represent strong evidence linking African and Eurasian hominoids.

Besides *Kenyapithecus*, another hominoid, *Otavipithecus*, is known from Namibia and dated to 13.0 ± 1.0 Ma (Conroy et al. 1992). Its phylogenetic position is questionable and is regarded by some as being closer to *Afropithecus* (Andrews 1992; Singleton 2000) and by others as more closely related to the
African apes and humans than are the Miocene apes of Eurasia (Pickford et al. 1997).

The Eurasian Middle Miocene hominoid fossil record is richer than the African one and includes the genera *Griphopithecus*, the recently found *Pierolapithecus* and *Dryopithecus*.

The genus *Griphopithecus* was originally found in the locality Devinska Nova Ves of Slovakia (Abel 1902), and later it was recognized in the localities Klein Hadesdorf, Austria (Steininger 1967), Engelswies, Germany (Heizmann 1992; Heizmann et al. 1996), Paşalar and Çandir, Turkey (Andrews et al. 1996). Its oldest appearance is that from Engelswies (Figure 1.3), dated to Late Orleanian/Early Astaracian (MN 5), while all the other localities are dated to Middle Astaracian, MN 6 (Mein 1999). Its synonymy problems with *Equatorius* have been reported above.

*Griphopithecus* is comparable in dental size to *Pan*, and has a robust mandible, strong superior and inferior transverse torus, very elongated and strongly inclined planum alveolar, and low-crowned molars with rounded cusps, thick enamel, and buccal cingulum. The few known fragmentary postcrania indicate an arboreal adaptation (Begun 2002; Kelley 2002). *Griphopithecus* fossils consist mainly of isolated teeth and are difficult to use as evidence for establishing phylogenetic relationships either to the African Middle Miocene hominoids or to the Late Miocene Eurasian ones. The limited maxillary and mandibular fragments from Paşalar preserve mainly primitive hominoid characters (Andrews et al. 1996). Its relationships with the Late Miocene Eurasian hominoids are not clear.

Recently, a partial skeleton of a hominoid was found in Barranc de Can Vila 1, Spain, named *Pierolapithecus catalaunicus*; it is dated to Middle Miocene or at 13.0–12.5 Ma. The main characters of *Pierolapithecus* are: low face, flat nasals, posteriorly situated glabella, highly originated zygomatic arches situated anteriorly at the level of M1, high nasoalveolar clivus, reduced heteromorphy in the premolars, elongated molars, absence of cingulum, large, low-crowned, and compressed upper canine, strong rib curvature, large and robust clavicle, and short metacarpals and phalanges. The morphology of the thorax (wide and anteroposteriorly shallow), the lumbar vertebrae, and the wrists indicate an orthograde position and locomotion characteristic for extant apes and humans. The facial morphology preserves the main derived characters of the extant great apes. On the other hand, *Pierolapithecus* retains some primitive characters, such as the short phalanges, suggesting palmigrade features. The overall facial and dental structures of *Pierolapithecus* indicate that it is probably close to the last common ancestor of great apes and humans (Moyà-Solà et al. 2004).
Dryopithecus is well known from Western and Central Europe, where it has been recognized in various localities in Spain, France, Germany, Austria, and Hungary (Figure 1.2). A maxillary fragment from Georgia, originally attributed to Udabnopithecus has recently been reassigned to Dryopithecus (Gabunia et al. 2001). Dryopithecus is known from the upper part of Astaracian (end of Middle Miocene) through the end of Vallesian (Figure 1.3). According to Agusti et al. (2001), the oldest Dryopithecus is known from the localities of Sant Quirze (Spain) and St. Stephan (Austria) dated to 11.7 ± 0.5 Ma. The youngest remains of Dryopithecus in Europe are those from the locality of Salmendingen (Germany), dated to the uppermost Vallesian. Dryopithecus from Udabno (Georgia) is also dated to the uppermost Vallesian (Figure 1.3). The genus includes four different species D. brancoi, D. crusafonti, D. fontani, and D. laietanus.

Two skulls of D. brancoi are known from Rudabanya, Hungary (Figure 1.5) characterized by short face, elongated neurocranium, weak but distinct supraorbital torus, biconvex vertically directed, and relatively short premaxilla, narrow nasal aperture, high and broad root of the zygomatic arches situated above the mesial half of M2, shallow canine fossa, large interorbital distance, squared-rounded orbits, and reduced foramen incisivum with short incisive canal (Kordos and Begun 1997, 2001). The cranial capacity of the Rudabanya skulls estimated at mean 320 cm³ for RUD-77 and 305 cm³ for RUD-197–200 (Kordos and Begun 1998, 2001). Another skull of D. laietanus is known from Can Llobateres, Spain, which resembles that from Rudabanya, but also has some differences, such as the very thick and strongly expressed anterior temporal ridge, the very high root of

![Figure 1.5](image)

Dryopithecus brancoi fragmentary skull (RUD-77) from Rudabanya, Hungary. Drawn from the illustrations of Kordos (1988). (a) Frontal and (b) lateral view.
the zygomatic arches, the flatter zygomatic plane, and the presence of a deep and broad depression above glabella in the frontal bone (Begun 2002).

The cranial morphology of *Dryopithecus* is a mixture of primitive and derived characters. The large interorbital distance, the shallow or absent supraorbital sulcus, the large frontal sinus, the flat or convex facial profile, and the round and great foramen palatinum are primitive characters very close to proconsulids. On the contrary, the deep maxilla, the elongation of the premaxilla, the elongated orbits, the inflated glabella, the presence of a weak supraorbital torus, and the nasomaxillary area are derived characters shared with extant great apes, *Ouranopithecus*, *Sivapithecus*, and *Australopithecus*.

The dentition of *Dryopithecus* is known from several remains found in various localities including this hominoid. The upper central incisors are narrow and high crowned and the canines are relatively small. There is a clear sexual dimorphism in the canines; the female ones are low crowned and small. The dentition is thin enameled in all *Dryopithecus* species. The upper premolars have proportions similar to the extant great apes, but the cusps are asymmetric as in *Proconsul*. The upper molars have relatively deep high cusps, well-developed occlusal crests, and accessory cusps in the distal margin like in *Ouranopithecus* and *Australopithecus*. The cingulum is absent in the upper molars and in this feature *Dryopithecus* is more derived than *Proconsul*. The lower incisors are high and buccolingually compressed. The P₃ has a large honing facet as in *Proconsul*. Compared to *Proconsul*, the molars are small and preserve a faint buccal cingulum, suggesting similarities to this taxon. The dentition of *Dryopithecus* preserves more primitive than derived features. The strongly marked dimorphism in the canines and their buccolingual flattening resemble that of primitive Early Miocene hominoids (like *Proconsul*), *Ouranopithecus*, *Gorilla*, *Pan*, and *Australopithecus*, while it is different from *Pongo* and *Sivapithecus*. The asymmetry of the upper incisors is a character present in *Proconsul*, *Dryopithecus*, *Ouranopithecus*, and *Gorilla*. It is stronger in *Pongo* and *Sivapithecus* and reduced in *Pan* and *Australopithecus*. The non-shovel-shaped I² is also a primitive feature present in the Early Miocene hominoids, as well as in *Ouranopithecus*, *Sivapithecus*, and the extant great apes except *Pan*. Moreover, the broad molars, the presence of cingulum in the lower ones, the large M³, the relatively higher cusps are characters suggesting similarities to *Proconsul*. On the other hand, the relatively reduced canines, the absence or reduced cingulum, the reduced asymmetry in P³, the relatively elongated P₄, and the presence of accessory cusps in the distal margin of M³ are derived characters, shared with *Ouranopithecus*, *Australopithecus*, and extant great apes.

Postcrania of *Dryopithecus* are known from various localities. The best known are those from Can Llobateres, Spain, including a partial skeleton,
which is considered in association with the recovered skull (Moya-Sola and Köhler 1996). However, some authors doubt this association (Begun 2002). Some postcrania are also known from Rudabánya, Hungary (Begun 1993). A humerus without epiphyses is also known from St. Gaudens, France (Lartet 1856). The humerus is long with rounded shaft, a flat and broad distal end, and with well-developed coronoid and radial fossa. The ulna has an anteroposteriorly flat proximal shaft, thick and prominent coronoid process, large radial facet, and well-defined median keel. The phalanges are elongated with curved shafts (Begun 1993, 2002). According to Moya-Sola and Köhler (1996), the morphology of the lumbar and thoracic vertebrae of Dryopithecus suggests orthograde postures.

The postcranial skeleton of Dryopithecus preserves some primitive hominoid features, like the slightly elongated lumbar vertebrae, the short metacarpals, and the small articular facets of the metacarpals and phalanges. On the other hand, it shares some derived characters, such as the very long forearm and the relatively short legs relative to its body size, the elongated and strongly curved proximal phalanges, the anteroposteriorly compressed shaft of the femur and the large clavicle, with Pongo, suggesting a relationship to the Pongo clade (Moya-Sola and Köhler 1996). The same authors believe that there was a common ancestor of Eurasian and African great apes, which was an orthograde climber. One of the Eurasian clades was adapted to suspension and slow climbing (present in Dryopithecus), while the African clade to more terrestrial quadrupedalism (Pan, Gorilla) or bipedalism (Australopithecus, Homo).

Dryopithecus is a Middle–Late Miocene hominoid, known from Eurasia. It preserves several primitive characters of the Early Miocene hominoids but also shares several derived characters with the younger hominids and recent great apes, linking the Early Miocene hominoids with Late Miocene Eurasian ones. Dryopithecus was dispersed in the whole of Eurasia and gave rise to several Late Miocene hominoids (Oreopithecus, Ouranopithecus, Ankarapithecus, Sivapithecus, and Lufengpithecus). According to Begun (2001), a member of the Dryopithecus clade entered Africa again and gave rise to the African apes and humans.

1.3.3 Late Miocene

African Late Miocene hominoids are rare represented by few remains, the majority of which were unearthed recently. A maxillary fragment of a large-sized hominoid is known from Kenya under the name Samburupithecus, dated to 9.5 Ma (Ishida and Pickford 1997). Although it has some similarities to Ouranopithecus, such as the low origin of the zygomatic arches and the thick enamel, its relationships with the extinct and extant hominoids are unclear. Recently, some
mandibular and postcranial remains of a hominoid named *Orrorin* have been found in Kenya; the locality is dated at \(~6.0\) Ma (Senut et al. 2001). *Orrorin* is characterized by jugal teeth that are smaller than those of australopiths, a small dentition relative to body size, large and not shovel-like \(I^1\), short \(C^\circ\), relatively deep mandibular corpus, and small \(m_2\) and \(m_3\) with thick enamel. The femur has a spherical head rotated anteriorly, an elongated neck with oval section and a mesially salient lesser trochanter. The humerus has a vertical brachioradial crest and the proximal manual phalanx is curved (Senut et al. 2001). *Orrorin* preserves some primitive ape characters, (deep mandibular ramus, anterior teeth, and \(p_4\)), as well as some more hominid ones (postcanine megadontia, postcranial evidences of bipedalism). According to Pickford et al. (2002), the femur of *Orrorin* shares some derived features with australopiths and *Homo*, but none with *Pan* or *Gorilla*, and among the Hominidae, it is closer to *Homo* than to australopiths.

In 2001, a skull with some mandibular and dental remains of a Late Miocene hominoid named *Sahelanthropus tchadensis* were found in Tchad; the associated mammal fauna indicates latest Miocene age, 6.0–7.0 Ma (Brunet et al. 2002). It is characterized by weak prognathism, small braincase, long and narrow basicranium, large canine fossa, small and narrow U-shaped dental arch, very wide interorbital distance, thick and continuous supraorbital torus, relatively small incisors and canines, rounded cusps in the molars, and moderate enamel thickness (Brunet et al. 2002). *Sahelanthropus* exhibits several primitive features, like the small braincase, the morphology of the basioccipital bone, and the position of the petrous portion of the temporal bone. But, it also preserves a set of derived features, such as small apically worn canines, medium-thick enamel, horizontally oriented and anteriorly situated foramen magnum, reduced prognathism, large and continuous supraorbital torus, and absence of canine diastema, which indicate close relationship to the hominid clade. On the basis of the presence of this mosaic of characters and the age of *Sahelanthropus*, Brunet et al. (2002) suggested that it belongs to the hominid clade and is close to the common ancestor of *Homo* and chimpanzees. Contrary to this opinion, Wolpoff et al. (2002) considered that the dental facial and cranial characters of this skull cannot define its position among hominids.

The sparse material of both *Orrorin* and *Sahelanthropus* and the nonfinal study of *Sahelanthropus* do not allow a clear definition of their phylogenetic relationships with the other hominids. Given that there had been no evidence for the presence of hominoids in the time interval from \(~8.5\) to \(4.5\) Ma until now, these two Late Miocene African hominoids constitute a link between the Miocene and Plio-Pleistocene hominids.

During Late Miocene, Eurasia was the evolutionary center of the hominoids with the presence of several taxa. Besides *Dryopithecus*, which appeared in Middle
Miocene but was still living in Late Miocene, *Oreopithecus*, *Ouranopithecus*, *Ankarapithecus*, *Sivapithecus*, *Lufengpithecus*, and *Gigantopithecus* are also known.

*Ouranopithecus* was originally recognized in the lower Axios valley, Macedonia, Greece (Bonis et al. 1973, 1990) and was later found in the Chalkidiki peninsula, Greece (Koufos et al. 1991). The *Ouranopithecus*-bearing mammal localities are: Ravin de la Pluie (RPl), Xirochori-1 (XIR), and Nikiti-1 (NKT), all dated to Late Vallesian or to the European land mammal zone MN 10 (Bonis and Koufos 1999; Koufos 2003). The magnetostratigraphic record suggests an age \( \approx \)9.6 Ma for XIR and \( \approx \)9.3 Ma for RPl (Sen et al. 2000); the fauna of NKT and its comparison to that of RPl indicates a younger age, estimated between 9.3 and 8.7 Ma (Koufos 2003). The analysis and comparison of the fauna indicate an open environment with shrubs, bushes, and trees across the banks of the rivers or around water spots (springs, lakes) (Bonis et al. 1992, 1999; Koufos 1980, 2006). The dental microwear of *Ouranopithecus* teeth indicates a hard-object eater, while that of the associated bovid fauna suggests that the grasses from the localities mainly consisted of graminae (Merceron et al. 2004). The known sample of *Ouranopithecus* includes maxillary and mandibular remains and a partial skull (Figure 1.6) with the face and the maxilla, but no postcrania, except for two manual phalanges.

*Ouranopithecus* has a large body size, in the range of variation of gorillas. There is strong sexual dimorphism expressed in the overall size of the dentition and in the size and morphology of the canines. The morphological characters of *Ouranopithecus* (Figure 1.6) are briefly as follows: well-developed supraorbital torus with a small glabellar depression, large interorbital distance, relatively short nasals, small, low, and quadrangular orbits, primitive nasoalveolar area with a large fossa incisiva, strong and low-rooted zygomatic arches, narrow and convex mandibular condyle, low and thick horizontal ramus, long planum alveolare with well-developed fossa genioglossa, strong superior and inferior torus, shovel-like incisors, relatively reduced canines, relatively symmetric upper premolars, low cusps in the molars, more symmetric P3 without honing facet, thick enamel, and absent or very weak cingulum (Bonis and Melentis 1977, 1978, 1985; Bonis and Koufos 1993; Koufos 1993, 1995; Bonis et al. 1998; Koufos and Bonis 2004).

The phylogenetic position of *Ouranopithecus* among the hominids is very important as it shares several derived characters with *Australopithecus* and *Homo*. Some of these characters are discussed below.

The lateral outline of the upper face is a derived character, as in primitive Early/Middle Miocene hominoids it is more oblique, while in *Ouranopithecus*, it is more vertical. The angle glabella–nasion–prosthion is 120°, while in the fossil and extant apes it reaches 140°. In *Australopithecus africanus*, it is 115° and in *A. boisei* 150° because of the flattening of the face (Bonis and Koufos 1993).
The structure of the nasoalveolar area of *Ouranopithecus* belongs to the primitive “African pattern,” like *Dryopithecus* and gorilla. It is different from that of *Pongo* and *Sivapithecus*, while it is closer to *Australopithecus* (Bonis and Melentis 1987; Bonis and Koufos 1994). However, it is more derived than *Proconsul* and *Morotopithecus* and less derived than *Australopithecus afarensis* and recent chimpanzee.

The narrow mandibular condyle of *Ouranopithecus* is a more hominine-like character. In the apes, the mandibular condyle is slightly convex, while in *Homo* it is narrower and more convex, and in *A. afarensis* it is more apelike (Bonis and Koufos 1997, 2001). The narrowness of the mandibular condyle can be measured by the index cranial–caudal diameter × 100/medial–lateral diameter. In *Ouranopithecus* it is 44.4, in *A. robustus* 33.8, in modern *Homo* 41.4–45.7, in *Gorilla* 46–56, in *Pan* 47–53, and in *Pongo* 43.3–43.9 (Bonis and Koufos 1993). Both the narrowness and shape of the mandibular condyle are related to the form of the temporomandibular joint. The latter is large and relatively flat in apes, while in hominines its anterior part is cylindrical and followed by a mediolateral directed fossa (Picq 1990). The capitulum of *Ouranopithecus* mandible could correspond to the latter kind of temporomandibular joint.
The cingulum is absent or very weak in all cheek teeth of *Ouranopithecus*, and compared to the strong cingula of *Proconsul* and pliopithecids this is a derived character. The presence of accessory cusps (-ids) at the distal ends of the upper and lower third molars, like in *Australopithecus afarensis*, is a derived character for *Ouranopithecus*, and this is related to the function of the teeth. The enamel thickness relative to the body weight of *Ouranopithecus* can be compared to that of some australopiths (*Paranthropus, Australopithecus africanus*), but it is thicker than *Dryopithecus, Sivapithecus, Proconsul africanus, P. major, Gorilla, Pan*, and *Pongo* (Bonis and Koufos 2001). Thus, enamel thickness could be a primitive feature relative to the recent hominoids but derived comparatively to the Early/Middle Miocene hominoids. The enamel thickness depends also upon the hardness of the food. A change to a more open environment, woodland, or wooded savannah, is associated with harder food items. Such an environmental change happened at the end of Middle Miocene in the Eastern Mediterranean leading to more open habitats (Koufos 2006) and to thick enamel in hominoids like *Ouranopithecus* and *Ankarapithecus*. On the other hand, in the Western Mediterranean the environment continued to be more forested and the hominoids are thin enameled (*Dryopithecus*).

The morphology of the P₃ in *Ouranopithecus* (not elongated, rounded occlusal outline, more symmetric buccal face, absence of honing facet, less oblique protocristid) distinguishes it from the apes. The wear pattern of the crown in P₃ is similar to that of *Australopithecus afarensis* as a result of the absence of the honing facet in both forms and the similar function of the tooth (Bonis and Koufos 2001). Moreover, the weaz pattern of the whole dentition of ouranopithecus is similar to that of A. afarensis (Koufos & Bonis, 2006).

The size of the upper canine of *Ouranopithecus* is a derived character. The plesiomorphic characters of the canine are the high, buccolingually flattened crown and the presence of a sharp distal crest on the buccolingual surface. This feature is present in cercopithecids, as well as in extant and Early/Middle Miocene hominoids. The shape of the canine is more rounded and in this feature it is very close to *Australopithecus afarensis*. The height of the upper canine of *Ouranopithecus* is relatively reduced (Bonis and Koufos 1993), while its morphology is completely different from that of the primitive hominids. On the other hand, the height of the crown is less reduced than in *Australopithecus afarensis* (Johanson et al. 1982). Comparatively the canine of *Proconsul* is three times larger than that of *Ouranopithecus*. The size of the canine compared to the size of the cheek teeth is also small in *Ouranopithecus* and similar to that of *Australopithecus afarensis* and female *Gorilla*. On the other hand, it is much larger (more than 120) in Early/Middle Miocene and extant hominoids (Bonis and Koufos 1993).
Concerning the lower deciduous dentition of *Ouranopithecus*, there are some derived characters shared with the Plio-Pleistocene hominines. The deciduous canine of *Ouranopithecus* is more reduced, compared to the molars, than that of *Proconsul* and *Ardipithecus* but less reduced than that of australopiths and *Homo*. The lower deciduous premolars of *Ouranopithecus* are more derived than Early–Middle Miocene and recent hominoids and less derived than australopiths and *Homo* (Koufos and Bonis 2004).

All these characters suggest that *Ouranopithecus* has strong relationships to the Plio-Pleistocene hominines as it shares several derived characters with them and can be considered as their ancestor. A cladistic analysis of 22 derived characters of *Ouranopithecus* shared with the extinct and extant hominoids, as well as with *Australopithecus* and *Homo* suggests that: (a) *Ouranopithecus* can be included in the subfamily Homininae, (b) the splitting of *Homo* and African apes is dated at more than 9.5 Ma, and (c) *Ouranopithecus* can be considered a sister group to *Australopithecus* and *Homo* (Bonis and Koufos 1997).

*Ouranopithecus* has several characters similar to those of australopiths; some of them are plesiomorphic (shape of symphysis, large interorbital distance, shallow or absent supratoral sulcus, large M³, development of nasomaxillary area). On the other hand, it preserves a set of apomorphic features. Some of them, such as thick-enamed teeth or canine reduction, may be homoplasies, but it would be peculiar and exceptional for all of them to be homoplasies (Bonis and Koufos 1999). There are some different opinions concerning this hypothesis. Begun (1992) suggests a relationship between *Gorilla* and *Ouranopithecus* and considers that *Dryopithecus* has slightly more evidence (gnathic structures and positional behavior) for being ancestral to African apes and humans. But, the postcranial morphology and consequently the positional behavior of *Ouranopithecus* is unknown. The two available phalanges of *Ouranopithecus* are different from all arboreal primates and closer to terrestrial forms. Begun (2002) agrees that *Ouranopithecus* and *Dryopithecus* “... comes closer to the ancestral morphology of the African apes and humans.” Earlier Andrews et al. (1996) made the statement that *Ouranopithecus* “is recognized as a hominine, related to the African ape and human clade and possibly close to the ancestry of the living species of this group.” In their phylogeny of the Hominoidea, *Ouranopithecus* is the possible ancestor of gorilla, chimpanzee, and humans (Andrews et al. 1996: fig. 12-7). Benefit and McCrossin (1995) considered *Ouranopithecus* to be linked with the African ape and human clade because of the presence of a supraorbital torus and the rectangular orbits. They also link *Samburupithecus* with *Ouranopithecus*, considering that it may be a potential candidate for membership in the African ape and human clade. However, the dental morphology of *Samburupithecus* differs from *Ouranopithecus* in having
more voluminous and higher cusps. But Africa is a huge region and who knows what will be discovered in the future. Recently a hominoid maxilla was found in the locality Çorak-Yerler (Turkey), which is referred to the Turolian. This hominoid is close to Ouranopithecus in the dental and palatal morphology, but it also has affinities to the early australopiths (Sevim et al. 2001; Begun et al. 2003). If this is true, it confirms the opinion that Ouranopithecus is ancestral to the australopith and Homo clade. The complete study and comparison of this specimen will provide interesting data about the phylogenetic relationships of the Late Miocene Eurasian hominoids and the Plio-Pleistocene hominids.

Another Asian Late Miocene hominoid is Ankarapithecus, found at Sinap-Tepe (Turkey), and known from three cranial remains. This hominoid is dated to ~10.0 Ma (Kappelman et al. 2003) and seems to be slightly older than Ouranopithecus. It was interpreted as having relationships with hominids (Alpagut et al. 1996), but recent analysis indicates phylogenetic relationships with the Sivapithecus–Pongo clade (massive maxilla, broad nasal aperture, high-placed zygomatic arches, narrow interorbital distance, very elongated nasal bones, absence of real supraorbital tori) (Begun and Güleç 1998). Thus, Ankarapithecus seems to be closely related to Sivapithecus and recent Pongo or to the Asian hominoids.

Oreopithecus is known from Italy only, where it is found at the localities of Bacinello, Casteani, Ribolla, Montebamboli, and Fiume Santo and is always associated with an endemic fauna (Rook et al. 1999). The localities have been dated to Middle–Late Turolian (Late Miocene) or at about 6.0–7.0 Ma (Harrison and Rook 1997; Steininger 1999). Oreopithecus is known by numerous cranial and postcranial remains including a complete skeleton. It can be considered as the best represented European Late Miocene hominoid.

The very small braincase, low zygomatic root, short and gracile premaxilla, long and narrow palate, narrow nasal cavity, projecting midface, and relatively high canines of Oreopithecus are primitive features distinguishing it from all great apes (Begun et al. 1997; Harrison and Rook 1997). The postcranial morphology of Oreopithecus resembles that of hylobatids and is related to suspensory positional behavior. It is also similar to that of the hominids and is related to large body mass in suspensory quadrupeds, with powerful grasping and high joint mobility (Harrison and Rook 1997). The majority of the morphological characters of Oreopithecus indicate that it is the most primitive known great ape (Harrison 1986b; Harrison and Rook 1997). On the other hand, Oreopithecus is considered as a highly derived member of the clade including all Eurasian Late Miocene hominoids (Moya-Solà and Köhler 1995, 1997) and is situated at the base of the hominoid radiation representing ancestral hominid morphology (Harrison and Rook 1997; Begun 2002). However, it is necessary to keep in
mind that *Oreopithecus* belongs to an endemic fauna and that some of its characters may be secondary adaptations to the local conditions.

*Sivapithecus* is known from India and Pakistan from three species: *S. sivalensis*, *S. indicus*, and *S. parvada*. All are dated to Middle–Late Miocene, from 12.5 to 8.5 Ma. A cranium and associated mandible is known from the Potwar plateau, Pakistan. The cranial and dental characters of *Sivapithecus* (tall and narrow ovoid orbits, very narrow interorbital distance, and “Asian” type of nasomaxillary region) indicate close similarities to the recent *Pongo*, and it is considered as its ancestor.

The hominoid *Lufengpithecus* is known from the Late Miocene of China, and it is characterized by large size, squared orbits, large interorbital distance, broad glabella, relatively short nasoalveolar clivus, moderate superior transverse torus, robust and prominent inferior torus, gracile appearance of the canines, high-crowned incisors, labiolingually thick I1, and relatively thick-enamelled molars with a dense and complex pattern of occlusal crenulations. The majority of the characters of *Lufengpithecus* suggest a close relation to the *Pongo* clade, but more material is necessary for more certain conclusions (Kelley 2002).

Two species of *Gigantopithecus* are known: *G. giganteus* from the Late Miocene of Siwaliks of India and Pakistan, and *G. blacki* from the Pleistocene of Southern China and Vietnam. It is a very large hominoid with relatively small lower incisors, low-crowned canines in comparison to the size of the mandible, slightly molarized premolars, and large premolars relative to the molars. *Gigantopithecus* is also considered as a member of the *Pongo* clade (Kelley 2002). It is the only known Miocene hominoid that survived into the Pleistocene.

All the Asian Miocene hominoids (*Ankarapithecus*, *Sivapithecus*, *Lufengpithecus*, and *Gigantopithecus*) belong to the *Pongo* clade, having strong relationships to the extant orangutans. Although geographically European, *Grifopithecus* is known from the Middle Miocene of Asia Minor and *Dryopithecus* from the Vallesian of Georgia (Andrews et al. 1996; Gabunia et al. 2001). Both areas are very close to Europe and may represent migration routes of hominoids from Africa to Eurasia and vice versa.

### 1.4 Conclusions

The ancestor of the hominids and recent great apes is included in the stem of Miocene hominoids, the common stock of hominids and apes. Despite the high increase in the fossil record of the hominoids during the last decades, it is still quite poor and limited for certain results. From millions of hominoids that lived for a long time (~20 Ma), very few and fragmentary remains are available, based
on which the scientists are trying to complete the chain of hominids’ evolution link by link. The absence of some links makes their work more difficult, and even the discovery of a possible link causes more problems than it solves. On the basis of recent data and known material, one can conclude that:

1. Oligocene *Aegyptopithecus* could be the possible ancestor of the Miocene hominoids, as it shares some derived features with the early hominoids. But it is worth mentioning the large time gap (~10 Ma) between this taxon and the first proconsulids.

2. Among the Early Miocene hominoids, *Proconsul* is the most possible link connecting the hominoids to the modern great apes and humans. Despite the known controversies, it preserves a mosaic of features, indicating relations to the extant great apes and humans. Despite the great similarities of *Proconsul* to the hominid clade, it is necessary to remember here that several other hominoid taxa are known from Early Miocene that could correspond to the possible link. But their limited and fragmentary material prevents certain comparisons and results.

3. The Middle Miocene *Dryopithecus*, known from Europe, seems to be an important taxon, which could be a possible link. Other Middle Miocene European or African hominoids could also represent a link, but as most of the discoveries are new with little material, their relations are still being discussed.

4. Several hominoids are known from the Late Miocene of Eurasia, while some new discoveries from Africa add important data to our knowledge. Some of these taxa, such as *Sivapithecus, Ankarapithecus, Lufengpithecus*, are closely related to the modern orangutans. The overall cranial and dental morphology of *Ouranopithecus* seems to be closer to the hominids and possibly it is the Late Miocene link connecting the hominoids to the extant great apes and humans. The recent discovery of *Orrorin* and *Sahelanthropus* in Africa, as well as the Çorak Yerler (Turkey) Turolian hominoid, has shortened the time gap between the Late Vallesian *Ouranopithecus* and australopithecines. However, the similarities of *Sahelanthropus* and *Ouranopithecus* (Brunet et al. 2002) indicate close relations between them.

The question of human origin will preoccupy scientists for a long time. Each new discovery will add some data, but it will also raise new questions and discussions among scientists. Despite this, it is very important that scientists continue their effort to collect more material and study in detail all the fossil and modern hominoids in order to understand their relations better. The continuing
efforts and the development of new methods will bring us closer to the real answer to the question.

Acknowledgments

I wish to thank very much Prof. W. Henke, Prof. H. Rothe, and Prof. I. Tattersall, editors of the present Handbook of Paleoanthropology, who invited me to participate as an author. Many thanks are also due to Dr. I. Sylvestrou for her nice work on Figures 1.1 and 1.4–1.6. I also thank Dr. P. Andrews and Dr. D. Begun for reading the manuscripts and making useful comments.

References


hominoïde *Ouranopithecus* dans les dépôts du Miocène supérieur de Macédoine (Grèce). CR Acad Sci Paris 327(2): 141–146


2 Defining Hominidae

Jeffrey H. Schwartz

Abstract

I review the relevant skeletal evidence that would unite a group that includes extant humans and their fossil relatives to the exclusion of living hominoids. Practicality suggests this group be called Hominidae. Potential hominid postcranial synapomorphies include a distinct angle at L5-S1, a long pubic ramus, a superoinferiorly short ilium that is roundedly expanded posteriorly, some thickening in the region of an iliac (crest) tubercle, a well-developed and knoblike anterior inferior iliac spine that lies noticeably superior to and somewhat back over the superior acetabular rim, a deep greater sciatic notch, differential distribution of cortical bone of the femoral neck, a defined linea aspera, femoral condyles that are of unequal size and different orientations with an outwardly slanted femoral shaft, and a concave lateral condylar facet on the proximal tibia that is at the same level with the medial facet, with the two facets separated by well-developed tubercles. Derived dental features that might unite Hominidae also characterize an orangutan clade. Within Hominidae, various subclades can be justified, suggesting that the relationships of some specimens referred to genus Homo lie within a broad “australopith” clade.

2.1 Introduction

By the twenty-first century, one would think that paleoanthropology would have long ago left behind the legacy of Linnaeus’ (1735) ultravague and systematically useless definition of our species, Homo sapiens: nosce te ipsum (know thyself). Yet, in spite of the incredible number of discoveries of fossils attributed to our clade (i.e., the clade that includes H. sapiens but excludes apes) since the mid-to-late nineteenth century—when discovery of the Feldhofer Grotto Neanderthal and especially of the Spy Neanderthals undermined the notion that humans were not ante-Diluvian—the history of paleoanthropology contrasts with that of the centuries-old disciplines of vertebrate and invertebrate paleontology in its apparent eschewing of taxonomic and systematic rigor. Consequently, the task of defining Hominidae is not as straightforward as one might imagine it should be.
2.2 Historical background

In the first detailed attempt to support Linnaeus’ subsuming humans within a specific group of mammals, and, in particular, grouping humans with apes, Huxley (1863) turned not only to hard- and even some soft-tissue morphology but also to development. Through ontogenetic comparisons, he argued that if a monkey could be distinguished developmentally not only as a mammal but also as a particular mammal, so, too, could humans. Hence, if a monkey was a primate, so was a human. Huxley then turned to “man’s place in nature” within Primates. However, since his premise was that humans were most similar to gorillas, he formulated his comparisons mostly between gorillas and other apes and then between gorillas and monkeys in order to argue that if a morphological “gulf” existed between gorillas and monkeys but not between gorillas and other apes, then humans could also be similarly allied with the apes. But whereas Linnaeus claimed that there were essentially no morphological features that distinguished humans from apes (Schwartz 1999a), Huxley believed otherwise. Consequently, in spite of demonstrating that anatomies as distinctive as the human foot were basically comparable to the grasping feet of apes and monkeys, Huxley concluded that humans were still sufficiently unique that they should be kept apart from the great apes, which he relegated to their own taxonomic group. In hard-tissue morphology, Huxley remained as impressed by aspects of the postcranial skeleton as Aristotle had been of the human thumb and Blumenbach of the human pelvic girdle and foot.

With his emphasis on aspects of the human postcranium and dentition, Blumenbach can be regarded as the “father of paleoanthropology” inasmuch as the criteria he delineated for distinguishing humans from other animals eventually became central to the identification of fossils as being “hominid” (Schwartz 1999a). For in 1795, in On the Difference of Man from Other Animals, Blumenbach emphasized various aspects of “the external conformation of the human body” as paramount in defining H. sapiens: erect posture; broad and flat pelvis; two hands; nondivergent hallux; close-set and serially related teeth; and some aspects of mandibular morphology.

As for erect posture, Blumenbach argued that, in contrast to other animals, this stance was natural and specific to H. sapiens as seen, for instance, in humans’ ossifying the tarsals before the carpals. He claimed that only humans have a “true” pelvis because the broad and expanded ilia form a basin that cups the viscera. Like Aristotle, he regarded the human hand as special because of its long thumb. In addition, because of the uniqueness of the human foot with its nondivergent hallux, Blumenbach also believed that the possession of only two handlike structures—true hands—was significant. Blumenbach pointed to the
human dentition as being distinctive in having orthally implanted lower incisors, canines that are not long or separated from the incisors, and molars with rounded cusps. He characterized the human mandible as being quite short, bearing a prominent chin, and having a distinctive articulation with the skull (presumably referring to the depth of the articular fossa), which, he suggested, was correlated with human omnivory.

Although he recognized that humans differ from many mammals in lacking a distinct premaxilla, Blumenbach also, but mistakenly, believed that other primates lack this bone—a claim he then used to argue against separating humans from other primates taxonomically. Goethe (1820) made a similar argument, but he did so on the (incorrect) belief that he could delineate a separate premaxillary bone in humans.

In addition to “the external conformation of the body” and in keeping with concerns of philosophers of the eighteenth century, Blumenbach (as well as Goethe) also addressed the “internal conformation” of humans: i.e., the importance of reasoning (as well as other mental attributes) as a criterion by which to distinguish humans from other animals, including other primates. Although not stated in these terms, we might identify a focus on mental attributes as underlying the subsequent emphasis in paleoanthropology on the size and external morphology of the brain [features that also attracted the attention of the eighteenth-century naturalists Buffon and Bonaparte (Schwartz 1987)].

The import of Blumenbach’s criteria for distinguishing humans from other animals came into play in paleoanthropology with the discovery not, as one might have supposed, of the Feldhofer Grotto or Spy Neanderthals remains, but with the discovery of the *H. erectus* specimens from Trinil, Indonesia (Schwartz 1999a). This history is likely due to Huxley’s (1863) argument that the Feldhofer Grotto Neanderthal represented not a distinct species but an extinct human whose cranial features merely extended into the past the continuum of racial “brutishness” seen in the Australian Aborigine (Huxley’s candidate for the most primitive of living humans). And it was Huxley’s envisioning of a morphological transition from “brutish,” “archaic” humans (whether extinct or extant) into “modern-looking” (i.e., European) humans that set the stage for perceiving human evolution as a continuum (Schwartz 1999a).

The Trinil specimens perforce split up Blumenbach’s criteria for defining *H. sapiens*: While the femur provided evidence of bipedalism (Blumenbach’s “erect posture”), the skullcap portrayed an individual less than fully human in its brain (and thus in its mental capacities). This unexpected combination of human and less-than-human features prompted Dubois to assign his new species name for the Trinil specimens (*erectus*) first to the genus *Anthropopithecus* (the taxonomic alternative to *Pan*), and then to Haeckel’s proposed genus for a
hypothetical extinct, speechless human relative, *Pithecanthropus* (=ape-man) (Schwartz 1999a). The implication one could derive from the *Pithecanthropus erectus* remains was that erect posture and bipedalism evolved prior to expansion and elaboration of the brain.

While lending itself to Darwin’s (1871) suggestion of a smooth transition from a semiquadrupedal African ape to an erect bipedal human, this picture—bipedalism first, brain second—was seemingly contradicted during the early 1900s with the discovery at Piltdown, England of a large, thin-boned, and rounded humanlike cranium, an apelike partial mandible preserving two molars, and an apelike lower canine. Under the presumption that these specimens were associated, the scenario of human evolution was turned around: Early human relatives became human first in their brains and then in the rest of the body (as inferred from the mandible and teeth). That is, the brain enlarged prior to the attainment of fully erect posture and bipedal locomotion. It was not until the 1950s, when the Piltdown fraud was exposed, that this alternative notion of human evolution—brain first, body second—was rejected. Before then, however, the discovery of the Taung child and Dart’s (1925) description and, more importantly, his interpretation of this specimen continued the intellectual trajectory Blumenbach had begun. But Dart conceived his scenario in the context of Darwin’s biogeographic premise of finding fossil evidence in Africa of intermediate forms that provided evidence of a morphological continuum between African apes and humans.

As Dart (1925 p 196) summarized his overall impression of the preserved craniodental features of the Taung specimen, this individual represented “an extinct race of apes intermediate between living anthropoids and man.” Dart depicted specific features—such as the configurations of the brow, nasal bones, zygomatic regions, orbits, and upper and lower jaws as well as the inferred skull shape—as being of “delicate and humanoid character” (Dart 1925 p 196). Perhaps most central to his speculations, however, were the size and potential details of the preserved partial endocast and also of the forward position of the foramen magnum (as indicated by bone adherent to the endocast). Dart (1925 p 197) assumed the latter was proof of this “humanoid’s” erect posture and then made the following extrapolations:

- The improved poise of the head, and the better posture of the whole body framework which accompanied this alteration in the angle at which its dominant member was supported, is of great significance. It means that a greater reliance was being placed by this group upon the feet as organs of progression, and that the hands were being freed from their more primitive function of accessory organs of locomotion. Bipedal animals, their hands were assuming a higher evolutionary role not only as delicate tactual, examining organs which were
adding copiously to the animal’s knowledge of its physical environment, but also as instruments of the growing intelligence in carrying out more elaborate, purposeful, and skilled movements, and as organs of offence and defence. The latter is rendered the more probable, in view, first of their failure to develop massive canines and hideous features, and secondly, of the fact that even living baboons and anthropoid apes can and do use sticks and stones as implements and as weapons of offence.

With regard to the Taung child’s brain (as represented by the endocast), Dart suggested that, since it was already as large as a chimpanzee’s and almost as large as a gorilla’s, it would have continued to enlarge, following a humanlike growth curve. In addition, as in humans but in contrast to apes, the Taung child’s brain was high and rounded, somewhat expanded in the temporal region, and bore an apparent lunate sulcus that was situated well posteriorly and inferiorly. On the basis of its seemingly more humanlike than apelike features, Dart inferred that this “humanoid” was humanlike in its faculties of “associative memory and intelligent activity.” The expanded cerebral cortex (as indicated by the supposed position of the lunate sulcus) also suggested to Dart that, in contrast to apes, the Taung “humanoid” had experienced increased sensory stimulation, both via vision (because of the forward position of the approximated orbits) and tactile sensation (because erect posture and bipedality supposedly freed the hands from involvement in locomotion). But the Taung child’s brain was not sufficiently enlarged in the temporal region for Dart (1925 p 198) to think that it had reached the “necessary milestone in the acquisition of articulate speech.”

As far as Dart was concerned, the Taung child, the name bearer of his new genus and species, *Australopithecus africanus*, displaced both Piltdown’s *Eoanthropus dawsoni* and Trinil’s *Pithecanthropus erectus* as viable “links” between humans and their apelike ancestors. Indeed, in spite of Dart’s statement about this juvenile extinct human relative’s intermediacy between humans and apes, his interpretation of it more closely reflected Blumenbach’s criteria for distinguishing *H. sapiens*. In 1925, then, while there were three somewhat but yet significantly different scenarios for the emergence of the earliest direct human ancestors, they all embraced the notion of an evolutionary continuum that extended from an apelike precursor, through a still unknown series of intermediates, to the most modern looking of living humans.

In terms of the focus of this chapter—defining Hominidae—further discoveries of potential extinct human relatives are less relevant than are attempts to integrate these fossils into a systematic framework that had originally been based on living *H. sapiens* alone. Thus, after a decade-and-a-half of successful fossil hunting in the limestone caves of South Africa and in caves and deposits in
Europe and Asia, a proliferation of genus and species names, and debates over the relationships to accepted hominids of the African taxa (referred to here as the australopiths), Clark (1940) was compelled to review the available evidence in order to determine the hominid status of any australopith.

In addition to echoing Huxley and Darwin’s assumptions of there having been a transformation from ape to human, Clark (1955) based his conclusions on what he later called the “total morphological pattern.” That is, he decided whether a particular fossil was hominid, not in terms of derived features it might share with humans but, rather, in terms of whether, overall, it resembled humans more than great apes. As will become obvious, this approach complicated matters further because the great apes were considered a group from which features for comparison could be extracted differentially from any ape, even if they characterized only one; Gregory (1922) had also used the device of “pick and choose” in arguing for an African ape–human relationship (Schwartz 2005). The irony of Clark’s using this phenetic approach is that, in 1955, he also made one of the clearest arguments for distinguishing between primitive and derived characters when making phylogenetic comparisons.

Although Clark (1940 p 317) concluded that australopiths were “more human than simian” especially in their teeth, his comparisons with apes were essentially restricted to the African taxa—which was a biased comparison he continued to emphasize throughout subsequent publications on fossil hominids. If he had included orangutans in his studies, he would, or at least should, have been struck by the similarities between this hominoid and australopiths in many details of facial and dental morphology (see Schwartz 2004a; also Schwartz and Tattersall 2005). This latter omission aside, Clark’s efforts to define Hominidae were as useless as his definition of the order Primates [i.e., primates are characterized by their lack, not sharing of any particular derived feature/s (Clark 1959)].

Mayr’s (1950) influential article on fossil hominids did not clarify the situation. Rather, on the grounds that all hominids were effectively adaptively similar because they were bipedal, Mayr collapsed all named taxa into one genus, Homo. Then after claiming that, since living humans are so diverse and occupy all available econiches the same must have been true for all hominids (thus precluding the opportunity and prerequisite for speciation), he subdivided his genus Homo into three time-sequential species: transvaalensis, erectus, and sapiens. Even though, in light of a much enlarged human fossil record, Mayr (1963) later “accepted” Australopithecus and made a nod toward Paranthropus as side branches that went extinct without issue, his concession did not elucidate how one decided in the first place if a specimen was a hominid, especially in the absence of postcranial remains. This is, indeed, a problem. For while it may be
true that some prominent scholars (e.g., Clark, Mayr, Washburn) decided “that the most important single factor in the evolutionary emergence of the Hominidae as a separate and independent line of development was related to the specialized functions of erect bipedal locomotion” (Clark 1964 p 14), the preponderance in the human fossil record of often fragmentary skulls and jaws and isolated teeth makes impossible identifying a specimen as being “hominid” on the basis of anatomical features believed to be reflective of bipedal locomotion. The exception, of course, is the region of the foramen magnum and occipital condyles.

2.3 Toward a definition of Hominidae

The task of defining Hominidae is twofold. First, there is a taxonomic decision. How widely does the classificatory net, Hominidae, extend? To chimpanzees? Chimpanzees and gorillas? All great apes? Although Clark (1955, 1964) wrote at a time when the great apes were believed to constitute a group or clade, and were thus classified together in the family Pongidae, his rationale for recognizing a group or clade Hominidae is, I believe, still viable and useful. Hominidae is a monophyletic group that subsumes extant humans and their fossil relatives to the exclusion of any living relative.

Some scholars who embrace molecular similarity between humans and chimpanzees as indicating evolutionary closeness then also group these hominoids in the same genus on the grounds that this endeavor is “cladistic” (Goldberg et al. 2003). Unfortunately, this taxonomic device hinders attempts to make sense of a clade (no matter what it is called) that subsumes humans and their closest relatives. For our closest relatives are not extant apes. They are as yet an unknown number of now extinct taxa. Furthermore, Goodman’s and like-minded molecular systematist’s claims to being “cladistic” in folding Pan and all australopiths into the genus Homo is relevant only in the context of translating a presumed scheme of phylogenetic relationship into a classification, as Hennig (1966) advocated. In truth, however, the asserted correlation between molecular identity and evolutionary propinquity is only an assumption of continual molecular change (Zuckerkandl and Pauling 1962; Caccone and Powell 1989). Furthermore, the aura of being “cladistic” is an artifact either of embracing this assumption and then stating that greater molecular similarity must therefore reflect a greater degree of shared derivedness or of rooting a tree in a taxon that is from the outset presumed to be primitive relative to those taxa of phylogenetic interest and then asserting that the similarities between the latter are synapomorph (see discussion in Schwartz 1987, 2005a, 2005b). In reality, the only constant source of mutation is UV radiation, whose effects are not only very
infrequent (typically $10^{-8}$ to $10^{-9}$) but also random with regard to impacting somatic versus sex cells (the latter being the only cells relevant to questions of relatedness).

Even though the notion of DNA being the “blueprint of life” has long outlived its utility, there still seems to be a lingering intuitive “sense” that molecular similarity should be phylogenetically revealing and also somehow be reflective of morphological similarity. But in metazoans, RNA delineates different reading frames from the same stretch of DNA, and it and various classes of proteins orchestrate signal transduction pathways and maintain cellular and molecular homeostasis (Carroll 2003; Maresca and Schwartz 2006; Schwartz 2005a, b). Indeed, the very nature of a cell’s biology to edit out or at the very least to suppress potential “mutations” both calls into question the “molecular assumption” of continual and accumulative mutation (cf. Caccone and Powell 1989) and more properly directs attention to the molecular and cellular (including physical and epigenetic) interactions that ultimately give rise to a functioning organism (Schwartz 2005). Consequently, from a molecularly based developmental perspective, morphology is indeed a viable resource in determining phylogenetic relationships (Schwartz 1999b, 2005b).

The latter point is not trivial because, in various cases, including that of *H. sapiens*, most potential insight into the evolutionary history of a clade comes from the fossil record. And, of course, fossils represent the physical remains of organisms—the results of their developmental histories. It is, therefore, odd that paleoanthropologists who accept without question a close relationship between humans and chimpanzees—for which there is virtually no morphological support (cf. Schwartz 1988, 2005; Shoshani et al. 1996; contra Begun 1994; Begun et al. 1997), whose claim for demonstrating a human–chimpanzee on the basis of morphology is merely an extrapolation from an assumed relationship between “*Australopithecus*” (undefined) and *Pan* (cf. Conroy 1994)—and, consequently, reject morphology as being phylogenetically revealing can then turn to morphology in order to analyze fossils (Pilbeam 1986; Asfaw et al. 1999).

Nevertheless, this contradiction leads to the second issue in defining Hominidae: the criteria one uses. If we accept the “molecular assumption” and its consequences—that morphology, especially fossilizable morphology, is phylogenetically unrevealing (Collard and Wood 2000; Pilbeam 2000)—then we are left with no alternative other than to throw up our hands in despair and wonder in amazement at how similar in various distinctive ways the morphologies of what we have been identifying as fossil hominids truly are. Hominidae as a potential clade is reduced to the single genus and species Blumenbach sought to define: *H. sapiens*. Nothing that has been considered an extinct hominid can be accepted
as such. Of course, when put in these terms, this is a ludicrous suggestion. Nevertheless, it is the logical extension of the “molecular assumption.”

As Carroll (2003) pointed out while warning against a naive expectation of developmental and evolutionary “insight” when human and chimpanzee genomes are ultimately compared in detail, there is not a one-to-one correlation between DNA sequence and morphology. An organism arises as a result of the ways in which segments of DNA are differentially translated and introns differentially spliced and the resultant products and other molecules are recruited and deployed. Lovejoy et al. (1999) offered a compelling argument in this regard, one that is not unrelated to the task of defining Hominidae. As they suggested (p 13428) with regard to the emergence of the clearly derived human pelvic form, “if a particular PI [positional information] gradient were to span \( n \) cell diameters, and those cells defined the ultimate anteroposterior dimension of the presumptive ilium (superoinferior in the adult human), then a slight increase in the steepness of its slope would cause that signal to span fewer cells, ‘distorting’ the presumptive anlagen and substantially altering downstream adult morphology.” That is, although received wisdom may embrace a process of change involving myriad small steps, “the transformation of the common ancestral pelvis [in its entirety] into that of early hominids may have been as ‘simple’ as a slight modification of a gradient” (Lovejoy et al. 1999; comments added). Consequently, while we might not know every aspect of molecular signaling and cell interaction underlying the emergence of a complete organism, we can still hypothesize from a grounding in cell biology that similar morphologies result from similar molecular and cellular interactions (Schwartz 2005b). As such, we can continue to test our theories of homology as we have all along in systematics: with other theories of homology.

Lovejoy et al.’s (1999) developmentally informed suggestion also brings with it other considerations relevant to the task of defining Hominidae (or any clade). What we might correctly interpret as major morphological change is not necessarily due to incremental and cumulative transformation over a significant period of time but may result from an integrated cascade of change that produces large-scale novelty rapidly, not slowly. This, of course, is theoretically expected (Schwartz 1999a), especially given the cell’s inherent mechanisms for maintaining homeostasis and resisting change (Maresca and Schwartz 2006; Schwartz 2005b). That is, it is likely that extremes of cellular perturbation, and not a history of accumulated “genetic change” (whatever that actually means), may permit the expression of novel avenues of molecular communication and cellular interaction that are normally suppressed or eliminated (Maresca and Schwartz 2006).
Recognizing that the disruption of DNA and cellular homeostasis can lead to large-scale morphological change that is abruptly expressed [in multiple individuals (Schwartz 1999a)], and which, if not lethal, may persist (Maresca and Schwartz 2006; Schwartz 1999a), makes futile the search for evidence—fossil or otherwise—of a smooth, incrementally continuous transition from one character state to another or from one taxon to another (Schwartz 1999a). In addition, even a superficial appreciation of the interplay between molecular communication, cell interaction, and organismal development highlights the problems of infusing systematic methodology with expectations of “discovering” evidence of transformation series. I refer here particularly to the often-employed concepts of “incipient” and “vestiges of”: “incipient” implying that a feature is “on the way” to becoming something else or at least a different version of itself (e.g., an incipient supraorbital torus in the case of Dryopithecus being touted as an ancestor of African apes and hominids), and “vestiges of” reflecting the idea that, since morphological features evolved “in order to” perform a certain function, one might find hints of them in organisms that no longer engage or engaged in that activity, but whose ancestor did (e.g., a vestige of a facet on the distal ulna of some specimens attributed to Australopithecus afarensis that one finds occasionally in gorillas, wherein knuckle walking is taken as the ancestral condition for hominids).

With these caveats in mind, we might now turn to the matter of defining Hominidae but not from the perspective of looking for the “defining” moment in a transition from some apelike condition to something seemingly hominid (either by a subtle hint of a supposed hominid trait or traces of a presumably primitive and retained feature). Inevitably, we return to Blumenbach’s list of criteria, to which other features have subsequently been offered up as delineating a clade that I shall refer to as Hominidae: humans and their extinct relatives.

### 2.4 Defining characters of Hominidae?

#### 2.4.1 Traditionally accepted features of “erect posture”

As reviewed above, Blumenbach’s emphasis on “erect posture” and “two handedness”—or, as Clark (1964 p 14) put it, on “specialized functions of erect bipedal locomotion”—has loomed large in considerations of the constitution of our own clade. Pilbeam (1972 p 62), for example, summarized some of the “adaptations” that are apparently associated with these “specialized functions of erect bipedal locomotion”: a vertebral column with a distinct lumbar curve that is set at a sharp angle relative to the sacrum; a “carrying angle,” wherein the lateral femoral condyle is larger and more weight bearing than the medial one, and the shaft of
the femur angles up and laterally away from the knee joint; a nongrasping foot with short toes and nondivergent hallux through which weight is transmitted during locomotion; and metacarpals in which the heads contact the substrate while the distal ends are elevated to form a springlike, transverse arch. We might also include Blumenbach’s description of the pelvic region as being bowl shaped (i.e., broad and shallow) and having a short, potentially laterally flaring, posteriorly expanded, anteriorly truncated ilium, a somewhat forwardly oriented acetabulum (which is also reflected in the orientation of the proximal femur relative to the shaft), a defined greater sciatic notch, and a broad, short sacrum, wherein the alae are not remarkably small relative to the size of the lumbar facet (Clark 1964; Schultz 1968). Clearly, these features distinguish living H. sapiens from other extant primates. However, the degree to which these characteristics are expressed in what have been identified as fossil hominids, and whether the appropriate postcranial remains are known, leaves much open for questioning.

For instance, among fossil specimens attributed to “anatomically modern” H. sapiens that probably represent this taxon (Schwartz and Tattersall 2000a, b, 2003) only Qafzeh 9 is known from a fairly complete, albeit, extremely crushed postcranium. Inasmuch as distortion of the skull and mandible compromises the definitive identification of a bipartite brow with a “glabellar butterfly” and a “true” chin with an inverted “T” configuration and thickened inferior symphyseal margin (Schwartz and Tattersall 2000a, b, 2003), postcranially, the pelvic region appears to have the details of H. sapiens (personal observation). Other specimens attributed to “anatomically modern” H. sapiens—especially Qafzeh 6 and all from Skhul—with sufficiently preserved crania and/or mandibles do not present a bipartite brow or an inverted mandibular symphyseal “T” (Schwartz and Tattersall 2000a, b, 2003). Known Skhul postcranial remains are incompletely representative and so crushed and poorly reconstructed that one can merely get a sense of their conforming to the above-mentioned pelvic configurations (personal observations).

Collectively, while Neanderthal postcrania (especially those associated with bipedal locomotion) differ from those of H. sapiens in details of size, shape, and/or morphology [e.g., more posteriorly expanded ilia, superoinferiorly tall and anteroposteriorly compressed pubic symphyseal region, relatively very elongate pubic ramus (and thus very wide subpubic angle), smoothly “hook-shaped” greater sciatic notch, smaller and differently oriented iliac auricular region, relatively large proximal and distal femoral ends, very large acetabulum, truncated calcaneus (Trinkaus and Howells 1979; Trinkaus 1983; Rak 1990; G. Sawyer reconstruction, personal observations)], they can still be accommodated by the general configurations summarized above. In addition, Neanderthals had a “carrying angle” and a lumbar curve.
With regard to Middle Pleistocene fossils, aside from the unique collection derived from Atapuerca’s Sima de los Huesos (Arsuaga et al. 1997), there are few instances where postcrania are associated with craniodental remains, Arago 44 (Day 1982; Sigmon 1982, personal observation) and the Jinniushan specimen (Rosenberg and Lu 1997) being obvious exceptions. With regard to the task of defining Hominidae, information on the pelvic region is the most readily available. Although differing from H. sapiens especially in some details of the ilium (e.g., flare, anterior superior iliac spine), Pilbeam’s criteria are still applicable to these fossils, as is also a femoral carrying angle in Sima specimens (Day 1986, personal observation; Rosenberg and Lu 1997).

Among earlier Pleistocene hominids, the Nariokotome (KNM-WT 15000) femora, vertebrae, and pelvic region, the Olduvai OH 8 foot bones, the Trinil and various other femora (especially Koobi Fora KNM-ER 1481), and the OH 28 and other os coxae (KNM-ER 1808 and 3228) are most relevant. Rose (1984), for example, compared the ER 1808 and 3228 favorably with those from Arago and Olduvai and also noted that they differ from H. sapiens and H. neanderthalensis in being relatively larger in the anterior region of the ilium. Although this region is not well preserved in either os coxa of KNM-WT 15000, it seems to compare well Arago 44, OH 28, and ER 1808 and 3228 (Day, 1986; Walker and Leakey 1993). An interesting but not especially emphasized aspect of these ilia is that, when the acetabula are oriented laterally, the “inner” surface of the blade faces noticeably forward (as in Sts 14), rather than medially as in H. sapiens and H. neanderthalensis (and also SK 50). The lumbar region of WT 15000 is somewhat curved, and there is a distinct angle at L5-S1 (cf. reconstruction in Walker and Leakey 1993). With regard to femora, WT 15000, ER 1481, and the complete Trinil specimen present a carrying angle, and in all the head is large. Interestingly, however, while the femoral neck is short and rounded in cross section in ER 1481 and Trinil, it is relatively long and somewhat flattened anteroposteriorly in both WT 15000 femora. The OH 8 foot bones, while reconstructed by Day and Napier (1964) with an arch and a first digit that was aligned with the other digits, may instead have had a semiopposable hallux (Clarke and Tobias 1995) (thereby suggesting that the proximal ends of the metatarsals were not elevated into an arch). If the OH 8 foot bones are those of a “hominid” (they were not associated with other skeletal or dental elements), then the pedal features in Pilbeam’s list do not define this potential clade.

Among Plio-Pleistocene specimens, the lumbar region and apparently also the lumbosacral articulation of Sts 14 essentially conform to Pilbeam’s criteria (Robinson 1972). In addition, the ilia of Sts 14 are both low and squat superoinferiorly, roundedly expanded posteriorly (and thus appear long anteroposteriorly) and have a well-defined greater sciatic notch. The preserved left os coxa
from the small Hadar AL 288-1 and the Makapansgat juvenile ilia (MLD 7 and 25) are similar to Sts 14 in these features, as apparently was also the right partial os coxa from Swartkrans (SK 50), which lacks the regions of the iliac crest and pubis (Robinson, 1972; Johanson et al. 1982; Lovejoy et al. 1982). With regard to adult ilia, at least Sts 14 and AL 288-1 and possibly also SK 50 differ from specimens discussed above in being oriented more laterally than vertically (thus exposing upward more of the internal surface of the ilium) and in having less blatantly “S”-shaped iliac crests; in addition, these specimens are only somewhat thickened in the region of an iliac (crest) tubercle and bear poorly developed iliac pillars (cf. Robinson 1972; Johanson et al. 1982; Day 1986). The preserved pubic rami of Sts 14 and AL 288-1 are reminiscent of those of Neanderthals and Jinniushan in being relatively long, but the symphyseal regions of the former two specimens are not also superoinferiorly tall and anteroposteriorly compressed (cf. Robinson 1972; Johanson et al. 1982; Day 1986; Rosenberg 1998). A superior view of the articulated Sts 14 pelvis illustrates (in contrast to H. sapiens) the relation of the elongate pubic rami to the relatively wide pelvic canal and the relatively posterior positioning of the outwardly flared ilia (cf. Robinson 1972). Furthermore, while in H. sapiens (and other Pleistocene specimens surveyed above) the curve of the iliac crest positions the anterior superior iliac spine just lateral to the parasagittal plane that intersects the posterior superior iliac spine, in Sts 14, the anterior portion of the ilium would have lain well lateral to the plane of the sacroiliac articulation (Robinson 1972). Even though only the left os coxa of AL 288-1 is known, it was likely similar to Sts 14 (cf. Johanson et al. 1982). In spite of the fact that the iliac crest of SK 80 is similar to Sts 14 in not being strongly “S-shaped,” when their ilial blades are oriented in the same plane, the acetabulum of SK 80 is similar to H. sapiens in facing laterally and slightly downward (Robinson 1972). In contrast, the fairly vertically aligned acetabulum of Sts 14 faces forward. When compared in the anatomical position, the Sts 14 ilium again assumes a more outwardly oblique orientation while SK 50 is more anteroposteriorly oriented, as in H. sapiens and various other specimens of “Homo.”

The anterior part of australopith ilia present a dichotomy of morphology. In the better preserved left os coxa of Sts 14, this region appears to be roundedly expanded anteriorly (Robinson 1972) as in StW 431 (Toussaint et al. 2003) and AL 288-1 (Johanson et al. 1982). Thus, a definitive anterior superior iliac spine cannot be identified. But in SK 50, even though its iliac crest is damaged along much of its length, what is preserved continues forward to become a well-defined, beak-shaped anterior superior iliac spine that projects markedly anterior to a bluntly thickened, almost knoblike anterior inferior iliac spine (cf. Day 1986; Robinson 1972). The smaller Makapansgat juvenile ilia (MLD 7 and 25) are similar to SK 50 (but not to
Sts 14 and AL 288-1) in having a projecting, beaklike anterior superior iliac spine (cf. Dart 1949, 1957; Robinson 1972). Although differing from SK 50 and MLD 7 and 25 in the region of the anterior superior iliac spine, Sts 14 and AL 288-1 are similar to them in developing a knoblike anterior inferior iliac spine that lies noticeably superior to and back over the superior margin of the acetabulum, as in other potential hominids surveyed above.

As is well known, australopiths have traditionally been interpreted as being postcranially intermediate between knuckle-walking great apes and “Homo”—as noted, for instance, in their developing a humanlike posterior iliac expansion while supposedly retaining an apelike anterior iliac distension. Given the above comparisons, this scenario seems inappropriate. Not all australopiths are characterized by the same configuration of the region of the anterior superior iliac spine. Only SK 50 and MLD 7 and 25 compare at all favorably with great apes in having a beaklike anterior superior iliac spine that continues forward the trajectory of the iliac crest. The rounded anterior expansion of this region in Sts 14 and AL 288-1, while absolutely and relatively large compared to H. sapiens and other possible fossil hominids, is, nevertheless, derived in its own right (Figure 2.1).

In contrast to hominoids in general, the acetabula of Sts 14, AL 288-1, and SK 50 are relatively and absolutely small (Robinson 1972; Day 1982; Johanson et al. 1982). Also in contrast to hominids in general, the head of the associated femur of AL 288-1 is relatively small and does not extend much circumferentially beyond the dimensions of the neck (especially in lateral view), and the femoral neck is

- Figure 2.1
Os coxae of Sterkfontein Sts 14 (cast) and H. sapiens. Although both have a knoblike anterior inferior iliac spine, its position differs, as do the orientations of the ilia and acetabula (which is relatively smaller in the extinct hominid). See text for further discussion; not to scale. (© J. H. Schwartz)
long and somewhat flattened or compressed anteroposteriorly (Lovejoy et al. 1982). Other preserved proximal femora (e.g., AL 128-1, 333-3 and 333-95, SK 82 and 97, BOU-VP-12/1, and KNM-ER 738 and 1503 and apparently also 1475) are essentially similar to AL 288-1 (cf. Napier 1964; Day 1986 (Figure 2.2). The Maka, Bouri (BOU-VP-12/1), and Sts 14 proximal femoral fragments lack the

head, but in preserved parts their necks are similar to the specimens just mentioned (cf. Robinson 1972; White et al. 1993; de Heinzelin et al. 1999). Since the broader comparison indicates that a long, compressed femoral neck is derived among hominoids, it may be phylogenetically significant that the femoral neck of the Nariokotome specimen (KNM-WT 15000) is elongate (Walker and Leakey 1993), as is also the neck of the proximal femur (BAR 1002'00) from the Late Miocene site Lukeino, Kenya that Pickford et al. (2002) referred to *Orrorin tugenensis* [the type specimen of which is an unassociated jaw with a few teeth (Senut et al. 2001)] (Figure 2.3).

Specimens with a “carrying angle” (characterized by the development of distal femoral condyles of different size and orientations and a laterally slanted femoral shaft) are known, for example, from Sterkfontein (Sts 34), Kromdraai (TM 1513), Hadar (AL 288-1), and Koobi Fora (e.g., KNM-ER 1475). The lateral slant of the femoral shaft is most pronounced in the former three specimens, but, in all, it is much more marked than in specimens or taxa discussed earlier.
To review, there are some features of the os coxa and femur that, in contrast to at least other catarrhines, appear to distinguish a subgroup of hominoids that we might refer to as “hominid”: i.e., a long pubic ramus, a superoinferiorly short ilium that is roundedly expanded posteriorly, some thickening in the region of an iliac (crest) tubercle, a well-developed and knoblike anterior inferior iliac spine that lies noticeably superior to and somewhat back over the superior acetabular rim, a deep greater sciatic notch, a defined linea aspera, femoral condyles that are of unequal size and different orientations with an outwardly slanted femoral shaft (producing a “carrying angle”), and a concave lateral tibial facet for the femur that is at the same level as the (also but primitively concave) medial facet, with the two facets being separated by well-developed tubercles.

Within this potential hominid clade, a subclade is distinguished by having an elongate and somewhat compressed femoral neck. The larger femoral head of specimens, such as WT 15000, ER 1547, and BAR 1002’00, is interesting in light of

---

1 Pickford et al. (2002) also argued that a greatly elongate femoral neck was a derived condition among hominoids but concluded that this was a shared derived state for Hominidae, not, as its more restricted distribution would suggest, a shared derived state of a less inclusive clade.
the possibility of their belonging to a clade united by femoral neck elongation: either these specimens and others like them are derived within the latter clade or, if their relationships lie with specimens attributed to “Homo,” femoral neck elongation is independently derived for them.

Before continuing, two features should be considered. One is the configuration of the distal femur when viewed from below (Figure 2.4). As initially illustrated by Robinson (1972), while TM 1513 and Sts 34 do not differ greatly from great apes in disparity and/or orientation of medial versus lateral condyles, the fossils are more strikingly different in being deeper anteroposteriorly and more trapezoidal (mediolaterally narrower anteriorly) rather than rectangular in outline. In further contrast to great apes, the patellar surfaces of TM 1513 and Sts 34 are somewhat more concave and slightly asymmetrical (i.e., in the anatomical position, the lateral margin of the patellar sulcus is more anteriorly elevated than the medial margin). The latter disparity is more obvious in TM 1513 than it is in Sts 34. The AL 288-1 distal femur is more similar to TM 1513 than to Sts 34 (cf. Lague 2002). In H. sapiens and H. neanderthalensis distal femora, for example, there is greater disparity between size and orientation of the medial versus lateral condyle, the bone is mediolaterally narrower and anteroposteriorly deeper, the patellar sulcus is deeper (and thus more roundedly “V” than broadly “U” shaped), and the lateral margin of this surface is much more noticeably protrusive anteriorly. As such, perhaps the features of the distal femur as particularly represented in Sts 34 had characterized the last common ancestor of a clade Hominidae, while the configuration of H. sapiens distinguished the ancestor of a subclade. Interestingly, the outline and specific configurations of the condyles and patellar sulcus as well as the margins of the preserved distal end of the left femur of WT 15000 more closely resemble australopiths (especially Sts 34) than H. sapiens.
The other feature to consider is the distance between the ischial tuberosity and the inferior acetabular lip, which is quite pronounced in great apes and apparently catarrhines in general (cf. Robinson 1972, personal observations; Aiello and Dean 1999). The separation between these two anatomical regions is marked in SK 50, shorter in Sts 34, and minimal in *H. sapiens*, in which a deep groove intervenes between the two structures. Interestingly, in WT 15000, the distance between these structures is reminiscent of Sts 34. Perhaps further study of this region will prove enlightening, if not in further defining features of Hominidae, perhaps in delineating a subclade within Hominidae.

### 2.4.2 More recently suggested postcranial features of “erect posture”

With the discovery at the Kenyan sites of Kanapoi and Allia Bay of specimens attributed to the species *Australopithecus anamensis*, other postcranial features were highlighted as distinguishing hominids from apes and, by extension, from other catarrhines (Leakey et al. 1995). Of particular note for this chapter is the configuration of the proximal tibia. In apes, the tibial facet for the medial femoral condyle is somewhat convex, while in specimens attributed to “accepted” australopith taxa and to species of *Homo* this facet is somewhat concave. In addition, the medial and lateral proximal tibial facets are separated by well-defined tubercles. One might also add to the tibial criteria for distinguishing between apes and potential hominids the fact that the proximal tibial facets are level with one another in “australopiths” and “*Homo*,” whereas in apes, the lateral facet lies below the level of the medial one (Aiello and Dean 1999; personal observations) (Figure 2.5).

![Figure 2.5](image)

**Figure 2.5**

Proximal tibiae, posterior view (from left to right) of *P. troglodytes*, Hadar AL 288-1A (cast), Kanapoi KNM-ER 29285 (cast), and *H. sapiens*. In the ape, note that the lateral condylar facet is convex anteroposteriorly and lies below the level of the shallowly concave medial condylar facet. In the hominids, the lateral condylar facet is concave and situated almost level with the medial condylar facet. See text for further discussion; not to scale. (© J. H. Schwartz)
In their description of the Lukeino proximal femur (BAR 1002′00), Pickford et al. (2002) commented on how slight the linea aspera is compared to specimens of australopiths and Homo (because the spiral line of BAR 1002′00 does not meet the linea aspera to form a structure of high relief). I agree (personal observation). They (p 202) then present “a precursor of the linea aspera” as a feature of bipedalism they would use to identify a femur as being hominid. Since, in BAR 1002′00, this faint muscle scar descends from the region of a gluteal tuberosity that lies below the greater trochanter, one can indeed distinguish this specimen, “australopiths,” and “Homo” from other hominoids. Apes do not present a muscle scar descending from the region of the gluteal tuberosity—which suggests that perhaps a “hint of a linea aspera” might define Hominidae.

Pickford et al. (2002) also mentioned that the lesser trochanter of BAR 1002′00 is medially projecting and, as in humans, gorillas, and Pongo, well separated from the femoral neck. Yet even among the illustrations in their article as well as those in Robinson (1972), Day (1986), and Walker and Leakey (1993), for instance, it appears that while a notable separation of lesser trochanter and femoral neck may characterize specimens such as SK 82 and 97, Sts 14, AL 288-1, KNM-WT 15000, and KNM-ER 738, 1503, and 1547, this configuration does not obtain to H. sapiens (or, for example, to Pan either). Perhaps this is due to differences in femoral neck length—which might imply that separation of the lesser trochanter from the femoral neck is phylogenetically significant within a clade Hominidae. The medial orientation of the lesser trochanter of BAR 1002′00, although similar to H. sapiens, is also characteristic of primates in general (cf. Swindler and Wood 1973; Aiello and Dean 1999; personal observation). In contrast, in specimens such as AL 128-1, 288-1, AL 333-95, AL 333-3, Maka VP 1/1, SK 82 and 97, Sts 14, WT 15000, and ER 738 and 1547, the lesser trochanter is posteriorly facing. Although not a defining character of Hominidae, perhaps the latter is a derived feature of a hominid subclade.

Among other features, Pickford et al. (2002; also Galik et al. 2004) suggested unite Orrorin via BAR 1002′00 with australopiths and at least H. sapiens, differential distribution of cortical bone of the femoral neck—thinner superiorly, thicker inferiorly—emerges as a potential defining characteristic of Hominidae (cf. Ohman et al. 1997).

2.4.3 Nonpostcranial features of “erect posture”

Although there are, no doubt, other features of the postcranium that might delineate a clade Hominidae, it might be prudent to turn now to another skeletal region from which “erect posture” or “bipedal locomotion” has been inferred: the base of the skull. For, ever since Dart’s (1925) discussion of the Taung child, a
forward position of the foramen magnum and its attendant occipital condyles has been central to the interpretation of australopiths and other early, potential hominids as bipedal hominoids. With *Ardipithecus ramidus* (White et al. 1994) and *Sahelanthropus tchadensis* (Brunet et al. 2002) being promoted as hominids, attention to the basicranium has never been more critical.

The interpretation of *Ardipithecus* (White et al. 1994) and *Sahelanthropus* (Brunet et al. 2002) as “hominid” was in part based on White et al.’s inference of bipedalism from the intersection of the bicarotid (foramen) cord and basion (on the anteriormost margin of the foramen magnum). The published photograph of *Sahelanthropus*’s crushed and distorted basicranium reveals, however, that what appears to be the anterior margin of the foramen magnum lies posterior to the bicarotid cord. In the undistorted state, the preserved left petrosal, and thus the bicarotid cord, would clearly have been more anteriorly placed.

Although not demonstrated, White et al.’s assertion of an association between basion and the bicarotid cord, an anteriorly placed foramen magnum, and erect posture and bipedalism seems intuitively likely. However, comparison of other potential fossil hominids, extant large-bodied hominoids, and various extant New and Old World monkeys, reveals a much more complex picture (Schwartz 2004a; Schwartz and Tattersall 2005; see also Wolpoff et al. 2002). In juvenile anthropoids, including humans, basion, the bicarotid cord, and the biporionic cord are essentially in alignment. This relationship is retained in the adults of some taxa, but, in others, the positions of basion and/or the bicarotid cord may change with growth relative to the biporionic cord. Consequently, it is not the bicarotid but the biporionic cord that appears to reflect the position of the foramen magnum. The alignment of basion and the two cords in the adult is, therefore, a neotenic feature, which, while not defining a clade Hominidae (Schwartz 2004a; Schwartz and Tattersall 2005), may be relevant to the delinea-
tion of relationships within it.

### 2.4.4 Proposed craniodental features of being hominid

Based on isolated teeth attributed to *Ardipithecus*, White et al. (1994) suggested that, in side profile, a hominid’s permanent upper canine has subequally long and quite divergent mesial and distal edges terminating in “shoulderlike” basal swellings that create the impression of a superoinferiorly short crown. This does not, however, describe the permanent upper canine (C\(^1\)) preserved in the skull of *Sahelanthropus* (Brunet et al. 2002) or of the majority of C’s of traditionally accepted Plio-Pleistocene and later hominids, including *H. sapiens* (cf. Schwartz and Tattersall 2005). Rather, White et al.’s description better captures the morphology of the C\(^1\)s of adult female orangutans and the deciduous upper canines of juvenile orangutans and chimpanzees (cf. Swarts 1988).
White et al. (1994) then described *Ardipithecus*’ C1 as “incisiform,” which derived from the belief, also shared by Brunet et al. (2002), that a C1-C1-P1 honing complex had been lost in hominids via a decrease in size and trenchantness of these teeth (particularly the canines) and closure of associated diastemata. According to the scenario, as the canines became less caniniform, they became increasingly associated functionally with the incisors, ultimately assuming the morphology and function of incisors. Yet, both the right C1 allocated to *Ardipithecus* and the C1 in the skull of *Sahelanthropus*, although differing in the triangularity of their buccal outlines, are apically pointed. Indeed, *Sahelanthropus*’ trenchant C1 would likely have occluded with a much more “caniniform”-looking C1 than the very unprimatelike isolated tooth thusly identified (Schwartz 2004a). Although it appears from teeth *in situ* that *Sahelanthropus* lacked upper diastemata, their absence is inferred for *Ardipithecus* from isolated teeth. Nevertheless, if *Ardipithecus* did not have diastemata, it and *Sahelanthropus* would be more derived than geologically younger (and supposedly descendant) specimens with diastemata [e.g., Hadar AL 200-1a and Sterkfontein StW 252 (Schwartz and Tattersall 2005); the maxilla from Bouri Hata allocated to *Australopithecus garhi* (photographs in Asfaw et al. 1999)]. Consequently, features associated with “reduction of a canine-premolar honing complex” do not define a clade Hominidae, but they might be delineate subclades within it.

Another approach to defining Hominidae is predicated on an *a priori* assumption of a human–chimpanzee relationship, which immediately constrains the hominid “outgroup” to a single large-bodied hominoid, i.e., *Pan*. Thus, although White et al. (1994 p 306) describe the dm1 of *Ardipithecus* as “apelike,” the only ape in their comparison was *Pan*. Perhaps not surprisingly, the dm1s of *Hylobates* and *Gorilla* are very similar to *Pan*. But *Pongo*’s differently configured dm1 is actually similar to its counterpart in traditionally accepted hominids, the major difference lying in the ape having more talonid cusp compression (cf. Schwartz 2004a; Swarts 1988). Otherwise, the dm1s of *Pongo*, “australopiths” and “*Homo*” conform to White et al.’s (1994 p 307) depiction of “apparently derived hominid features”: i.e., “buccolingual crown expansion, mesiolingually prominent metaconid, well-defined anterior fovea, and large talonid with well differentiated cusps” (White et al. 1994 p 307). In fact, the dm1s of *Pongo* and hominids2 are similarly derived compared to other extant hominoids in that the protoconid is more mesially situated; the less vertical and lingually facing anterior

---

2 The comparison includes juvenile “*Homo*,” including Melka Konturé MK81 GAR IV (2) (see Schwartz 1995; Schwartz and Tattersall 2002, 2003) and “australopiths” [e.g. Hadar AL 333-43b, Koobi Fora KMN-ER 820, 1477, and 1507, Laetoli LH2 and 3q, Omo 227, Taung, Swartkrans SK 47, Kromdraai TM1536, TM1601a, TM1604, KB5503, and KB5223, and Sterkfontein Sts24a (see Schwartz and Tattersall, in press)].
fovea (=trigonid basin) is noticeably smaller than the talonid basin and is enclosed by a distinct paracristid that courses somewhat mesially and then turns toward the lingual side; and the more horizontally oriented talonid basin is enclosed by a distinct hypocristid (personal observation; cf. Swarts 1988).

Only if *Pongo* (and other members of its clade for which dm1s are currently unknown, but for whom this description would be predicted as applicable) is not the sister taxon of a potential hominid clade would these dental features be relevant to defining the latter clade alone.

### 2.5 Additional defining characters of Hominidae?

The latter point is not trivial. For although some clearly derived features might in isolation delineate a clade Hominidae, they are also shared with *Pongo* and its potential extinct relatives. Consequently, in order to claim them as hominid apomorphies alone, and their presence in an orangutan clade as autapomorphy for it and nonsynapomorphic with hominids (i.e., homoplastic), one must first embrace another theory of extant large-bodied hominoid relationship and then “explain away” the phylogenetic significance of the similarities between humans and orangutans. It might therefore be useful to review briefly some of these features.

#### 2.5.1 *Pongo*-like dental and palatal features of potential hominids

An interesting case of what could be called “interpreter’s bias” lies in analyses of molar enamel thickness. For example, although Martin (1985) was cited as demonstrating that the last common ancestor of large-bodied hominoids had thick molar enamel (which was retained in orangutan and hominid clades but secondarily reduced in African apes), he in fact interpreted the enamel thickness data in the context of an (orangutan (human–African ape)) theory of relatedness (Schwartz 1987). Otherwise he should, or at least could, have concluded that thick molar enamel united human and orangutan clades. Pilbeam (1986; also Kelley and Pilbeam 1986) also interpreted molar enamel thickness in the context of an (orangutan (human–African ape)) theory of relatedness but presented two scenarios: Martin’s and one in which the last common ancestors of separate human and orangutan clades independently developed thick molar enamel. In the latter case, thick molar enamel would be a defining feature of Hominidae.
Most recently, Schwartz (2000) found that, while humans and orangutans both have thick molar enamel, humans have thicker enamel in some areas of their occlusal surfaces. He suggested that thick molar enamel had evolved independently in humans and orangutans (and, by implication, in the last common ancestors of separate human and orangutan clades) because of the demands of different diets. A more straightforward interpretation of Schwartz’s data is that human and orangutan clades shared a common ancestor that possessed thick molar enamel and that, within this clade, humans (and perhaps other hominids) are more derived. Only by accepting thick molar enamel as synapomorphic of traditionally accepted hominids can one embrace thin-enameled *Ardipithecus* as a sister of this clade (assuming that this relationship is based on synapomorphy). Of course, Brunet et al.’s (2002) suggestion that the thicker enameled *Sahelanthropus* is ancestral to *Ardipithecus* underscores the need for systematic rigor in paleoanthropology.

If thick molar enamel does not contribute to defining Hominidae, does dental morphology? Some aspects of it might, but, at present, delineating any would be a Herculean task. This difficulty derives from the longstanding notion of the Miocene being a time of an “ape radiation,” after which “hominids” emerged. Guided by this scenario, paleoanthropologists typically identified post-5.5 Ma specimens with thick-enameled cheek teeth as “hominid” and pre-5.5 Ma specimens as orangutan relatives. Nevertheless, study of fossils identified as hominid reveals that *Pongo*-like teeth have often been referred to species of “*Homo*” and “*Australopithecus*” (cf. Schwartz 2004b; Schwartz and Tattersall 1996; cf. Schwartz et al. 1995; Schwartz and Tattersall 1996, 2003; Schwartz 2004b; 2005a). Teeth allocated to *Orrorin tugenensis* (Senut et al. 2001), if truly thick-enameled, reinforce the question of how many Miocene “apes” are actually hominids and how many Plio-Pleistocene “hominids” are not. In addition, since “hominid” teeth are often worn flat, little attention has been paid to details of occlusal morphology. Thus, although most of the human fossil record has now been scrutinized at the level of morphs (Schwartz and Tattersall 2002, 2003, in press), much work remains in testing these hypotheses before assigning specimens (other than type specimens) to specific genera and species.

Were it not for their presence in *Pongo, Sivapithecus*, and *Ankarapithecus*, for instance, other features that could “define” Hominidae are the development of a single incisive foramen (Schwartz 1983, 1997) and a posteriorly thickened palate (as seen in midline cross section) (Schwartz 2004b). With regard to the latter feature, although it appears that the Hadar broken palate AL 200-1a thinned posteriorly, if true, this would be the outlier among “accepted” hominids (Schwartz 2004b; cf. Schwartz and Tattersall 2002, 2003, 2005). As with thick molar enamel and the morphological details of dm₁, the simplest phylogenetic
interpretation of these features is that they characterized the last common ancestor of human and orangutan clades.

2.5.2 *Pongo*-like features of “australopiths”: implications for defining Hominidae

Another stumbling block to defining Hominidae is the degree to which “australopiths” and specimens traditionally allocated to “Homo” differ from one another, especially in craniofacial morphology: e.g., in orbital and nasal aperture outline as well as the configuration of the supra- and infraorbital regions; elevation of the nasoalveolar clivus above the floor of the nasal cavity; height and orientation of the infranasal region; orientation, flatness, and height of the zygomatic region; the development of a “snout” with/without facial pillars (vertical or medially inclined) and a canine fossa; and in the development and configuration of a mastoid process. Thus, while one might unite all of these potential hominids via various postcranial features, craniofacial morphology seems only to delineate possible hominid subclades. This may in part explain why one approach to linking humans with *Pan* is to try first to link the latter with australopiths and then to assert that, since australopiths and *Homo* form a clade, humans and chimpanzees are closely related (Begun 1994, Begun et al. 1997). In the latter case, the primary feature of supposed synapomorphy between (an undefined) *Australopithecus* and *Pan*—barlike supraorbital torus with sulcus behind—describes no “australopith” and only some specimens of “*Homo*” (Kimbel 1986; Clarke 1987; Schwartz 1997; Schwartz and Tattersall 2002, 2003, 2005). In addition, the markedly inferosuperiorly tall supraorbital region of *Sahelanthropus* is not barlike, but it is so unusually tall that it must represent a derived, not primitive, configuration—which, together with its distinct dental morphology and not very thick molar enamel, makes determining its relationship to other hominoids difficult (Schwartz 2004a).

While seeking connections between humans and African apes (especially *Pan*) has historical precedent (Clark 1940; Johanson and White 1979), the most favorable comparisons are actually between *Pongo* (and its fossil relatives) and australopiths (Schwartz 2004b, 2005a): e.g., rimlike superior orbital margins with a rather smooth transition into the frontal plane; often ovoid orbits; and forwardly facing, tall, and often vertical zygomatic regions (a further derived configuration, e.g., as in KNM-WT 17000, is a posteriorly tilted zygoma). Many australopiths are also similar to *Pongo* in having distinct facial pillars that emerge just above the canines and, together with a variably developed canine fossa, delineate a “snout.” Interestingly, the Taung specimen has the long, slitlike
single incisive foramen seen in *Pongo* and also preserved in *Sivapithecus* and *Ankarapithecus* (Schwartz 2005). While I am not, however, advocating a special relationship between a *Pongo* clade and some or all australopiths, attention to these apparent synapomorphies is more relevant to the task of defining Hominidae, its subclades, and sister taxa than is currently appreciated.

### 2.6 Final comments

When comparing extant taxa, defining Hominidae is a simple matter. Through a curious historical twist, *H. sapiens* is the only survivor of a clade whose dimensions we still do not fully know. Take, for instance, the recent discovery of a possible hominid from the Late Pleistocene of Flores, Indonesia (Brown et al. 2004). Its combination of morphologies would be unexpected at any time period.

It has a moderately globular cranium (as in some hominids), somewhat thickened and anteriorly (but not superiorly) protruding but rimlike supraorbital margins with no sulcus behind (australopiths in part and orangutans and their relatives), tall, ovoid orbits (*Pongo, Sivapithecus, some “australopiths”), flat nasal bones (most hominoids, including some “*Homo*”), forwardly facing and vertical yet superoinferiorly short zygomas (australopiths and orangutans and their relatives), well-developed mastoid processes (some “australopiths” and some “*Homo*”), a thick frontal with thick diploe (autapomorphous or pathological), no frontal sinuses (most primates, including bonobos), an anterior cranial fossa that does not extend fully over the orbital cones (most primates, including some “*Homo*”), a clivus that slopes gently away from the dorsum sellae (most primates, including some “*Homo*”), basion, the bicornal and biporionic cords in alignment (juvenile anthropoids, some adults), a broad incisive foramen that proceeds anteriorly as a expanding groove (a few “australopiths”), marked separation of the nasoalveolar clivus and an anteriorly thin palate (African apes, some Miocene hominoids, some “australopiths”), a smoothly but narrowly curved mandibular symphyseal region (many anthropoids), a long retromolar space (various “*Homo*”), a broadly and smoothly rounded but somewhat truncated gonial angle (some “*Homo*”), a very anteroposteriorly long sigmoid notch (some “*Homo*”), a sigmoid notch crest that is deepest near the coronoid process (autapomorphic), very large cheek teeth and apparently relatively small anterior teeth (some “australopiths”), unusually mesiodistally short upper and lower molars with large mesial and truncated distal cusps (autapomorphic), a relatively long ilium (SK 50) with a beaklike anterior superior iliac spine (great apes, SK 50, MLD 7 and 25), a poorly defined iliac pillar (australopiths), a knoblike anterior
inferior iliac spine that lies noticeably superior to above and somewhat back over the supraacetabular rim (hominids as discussed here), an arcuate line that descends well before reaching the region of the acetabulum (some hominids), no ischial spine (most primates), posterior iliac expansion that defines a greater sciatic notch (hominids as used here), a “V”-shaped greater sciatic notch (most hominids, but not Neanderthals), a large femoral head (most primates), a long and anteroposteriorly compressed femoral neck (Ororrin “australopiths” including BOU-VP-12/1, WT 15000), a well-defined intertrochanteric crest (most primates, but not WT 15000, Ororrin, and most if not all “australopiths”), a medially facing lesser trochanter (most primates, not “australopiths”), a weakly developed linea aspera (Ororrin), a femoral “carrying angle” (“australopiths” and “Homo”), a tibia that is much shorter than the femur (at least apes), poorly differentiated tibial tubercles (most primates), a medial tibial facet for the femoral condyle lower than the lateral (at least apes), and a convex medial tibial facet (at least apes) (cf. text and illustrations in Brown et al. 2004).

Is this specimen—if the cranial and postcranial remains represent a single individual—a hominid? A gut reaction based on the external morphology of the skull is “yes,” but the internal morphologies are odd. The teeth are not necessarily hominid, the humeral shaft lacks the torque or “twist” characteristic of large-bodied hominoids (Morwood et al. 2005) and the morphology of the distal articular region is clearly not hominoid (cf. Schwartz 1986; Morwood et al. 2005), the tibia is definitely not hominid, the femur is Ororrin-like, and the partial os coxa is somewhat “australopith”-like. So why is the composite Flores specimen considered a species of Homo when its affinities to a clade Hominidae are not entirely clear? Largely because there has been a history of allocating specimens to taxa based more on their geological age than on their morphology—which, in turn, has led to the general practice of trying to explain away “anomalous” morphologies in terms of variation. Methodologically, however, before one even contemplates referring a specimen to a genus and species, one should have to defend first why any specimen is a hominid and then a member of the smaller subclades that subsume that species. But in order to do so, we must have a working definition of this potential clade that is open to criticism and revision. In this regard, there is still much work to be done.

**Acknowledgments**

Drs. Henke, Rothe, and Tattersall certainly provided a challenge by asking me to tackle this problem. In the end, it was an enlightening endeavor. But it would have
been impossible were it not for my long-term study with Ian Tattersall of much of the human fossil record, which depended on access to specimens (at the very least type specimens) afforded by most discoverers and museum curators in charge of specimens. Perhaps someday specimens from the Middle Awash, Ethiopia, and Chad will be generally available for study.

References

Blumenbach JF (1969) On the natural varieties of mankind (De Generis Humani Varietate Nativa) (1775 and 1795 volumes). Bergman, NY
Darwin C (1871) The Descent of Man, Parts I and II. John Murray, London
Defining Hominidae


Gregory WK (1922) The origin and evolution of the human dentition. Williams and Wilkins, Baltimore


Huxley TH (1863) Man's place in nature. D Appleton, NY


Linnaeus C (1735) Systema naturae. Leiden


Schwartz JH (1999a) Evolutionary provocations: Paul Sondaar, the evolution of the horse, and a new look at the origin of species. Deinsea 7: 287–300
Schwartz JH (2004a) Barking up the wrong ape: Australopiths and the quest for chimpanzee characters in hominid fossils. Coll Antropol 28(Suppl. 2): 87–101


2 Defining Hominidae

Jeffrey H. Schwartz

Abstract

I review the relevant skeletal evidence that would unite a group that includes extant humans and their fossil relatives to the exclusion of living hominoids. Practicality suggests this group be called Hominidae. Potential hominid postcranial synapomorphies include a distinct angle at L5-S1, a long pubic ramus, a superoinferiorly short ilium that is roundly expanded posteriorly, some thickening in the region of an iliac (crest) tubercle, a well-developed and knoblike anterior inferior iliac spine that lies noticeably superior to and somewhat back over the superior acetabular rim, a deep greater sciatic notch, differential distribution of cortical bone of the femoral neck, a defined linea aspera, femoral condyles that are of unequal size and different orientations with an outwardly slanted femoral shaft, and a concave lateral condylar facet on the proximal tibia that is at the same level with the medial facet, with the two facets separated by well-developed tubercles. Derived dental features that might unite Hominidae also characterize an orangutan clade. Within Hominidae, various subclades can be justified, suggesting that the relationships of some specimens referred to genus Homo lie within a broad “australopith” clade.

2.1 Introduction

By the twenty-first century, one would think that paleoanthropology would have long ago left behind the legacy of Linnaeus’ (1735) ultravague and systematically useless definition of our species, Homo sapiens: nosce te ipsum (know thyself). Yet, in spite of the incredible number of discoveries of fossils attributed to our clade (i.e., the clade that includes H. sapiens but excludes apes) since the mid-to-late nineteenth century—when discovery of the Feldhofer Grotto Neanderthal and especially of the Spy Neanderthals undermined the notion that humans were not ante-Diluvian—the history of paleoanthropology contrasts with that of the centuries-old disciplines of vertebrate and invertebrate paleontology in its apparent eschewing of taxonomic and systematic rigor. Consequently, the task of defining Hominidae is not as straightforward as one might imagine it should be.
2.2 Historical background

In the first detailed attempt to support Linnaeus’ subsuming humans within a specific group of mammals, and, in particular, grouping humans with apes, Huxley (1863) turned not only to hard- and even some soft-tissue morphology but also to development. Through ontogenetic comparisons, he argued that if a monkey could be distinguished developmentally not only as a mammal but also as a particular mammal, so, too, could humans. Hence, if a monkey was a primate, so was a human. Huxley then turned to “man’s place in nature” within Primates. However, since his premise was that humans were most similar to gorillas, he formulated his comparisons mostly between gorillas and other apes and then between gorillas and monkeys in order to argue that if a morphological “gulf” existed between gorillas and monkeys but not between gorillas and other apes, then humans could also be similarly allied with the apes. But whereas Linnaeus claimed that there were essentially no morphological features that distinguished humans from apes (Schwartz 1999a), Huxley believed otherwise. Consequently, in spite of demonstrating that anatomies as distinctive as the human foot were basically comparable to the grasping feet of apes and monkeys, Huxley concluded that humans were still sufficiently unique that they should be kept apart from the great apes, which he relegated to their own taxonomic group. In hard-tissue morphology, Huxley remained as impressed by aspects of the postcranial skeleton as Aristotle had been of the human thumb and Blumenbach of the human pelvic girdle and foot.

With his emphasis on aspects of the human postcranium and dentition, Blumenbach can be regarded as the “father of paleoanthropology” inasmuch as the criteria he delineated for distinguishing humans from other animals eventually became central to the identification of fossils as being “hominid” (Schwartz 1999a). For in 1795, in On the Difference of Man from Other Animals, Blumenbach emphasized various aspects of “the external conformation of the human body” as paramount in defining H. sapiens: erect posture; broad and flat pelvis; two hands; nondivergent hallux; close-set and serially related teeth; and some aspects of mandibular morphology.

As for erect posture, Blumenbach argued that, in contrast to other animals, this stance was natural and specific to H. sapiens as seen, for instance, in humans’ ossifying the tarsals before the carpals. He claimed that only humans have a “true” pelvis because the broad and expanded ilia form a basin that cups the viscera. Like Aristotle, he regarded the human hand as special because of its long thumb. In addition, because of the uniqueness of the human foot with its nondivergent hallux, Blumenbach also believed that the possession of only two handlike structures—true hands—was significant. Blumenbach pointed to the
human dentition as being distinctive in having orthally implanted lower incisors, canines that are not long or separated from the incisors, and molars with rounded cusps. He characterized the human mandible as being quite short, bearing a prominent chin, and having a distinctive articulation with the skull (presumably referring to the depth of the articular fossa), which, he suggested, was correlated with human omnivory.

Although he recognized that humans differ from many mammals in lacking a distinct premaxilla, Blumenbach also, but mistakenly, believed that other primates lack this bone—a claim he then used to argue against separating humans from other primates taxonomically. Goethe (1820) made a similar argument, but he did so on the (incorrect) belief that he could delineate a separate premaxillary bone in humans.

In addition to “the external conformation of the body” and in keeping with concerns of philosophers of the eighteenth century, Blumenbach (as well as Goethe) also addressed the “internal conformation” of humans: i.e., the importance of reasoning (as well as other mental attributes) as a criterion by which to distinguish humans from other animals, including other primates. Although not stated in these terms, we might identify a focus on mental attributes as underlying the subsequent emphasis in paleoanthropology on the size and external morphology of the brain [features that also attracted the attention of the eighteenth-century naturalists Buffon and Bonaparte (Schwartz 1987)].

The import of Blumenbach’s criteria for distinguishing humans from other animals came into play in paleoanthropology with the discovery not, as one might have supposed, of the Feldhofer Grotto or Spy Neanderthals remains, but with the discovery of the *H. erectus* specimens from Trinil, Indonesia (Schwartz 1999a). This history is likely due to Huxley’s (1863) argument that the Feldhofer Grotto Neanderthal represented not a distinct species but an extinct human whose cranial features merely extended into the past the continuum of racial “brutishness” seen in the Australian Aborigine (Huxley’s candidate for the most primitive of living humans). And it was Huxley’s envisioning of a morphological transition from “brutish,” “archaic” humans (whether extinct or extant) into “modern-looking” (i.e., European) humans that set the stage for perceiving human evolution as a continuum (Schwartz 1999a).

The Trinil specimens perforce split up Blumenbach’s criteria for defining *H. sapiens*: While the femur provided evidence of bipedalism (Blumenbach’s “erect posture”), the skullcap portrayed an individual less than fully human in its brain (and thus in its mental capacities). This unexpected combination of human and less-than-human features prompted Dubois to assign his new species name for the Trinil specimens (*erectus*) first to the genus *Anthropopithecus* (the taxonomic alternative to *Pan*), and then to Haeckel’s proposed genus for a
hypothetical extinct, speechless human relative, *Pithecanthropus* (=ape-man) (Schwartz 1999a). The implication one could derive from the *Pithecanthropus erectus* remains was that erect posture and bipedalism evolved prior to expansion and elaboration of the brain.

While lending itself to Darwin’s (1871) suggestion of a smooth transition from a semiquadrupedal African ape to an erect bipedal human, this picture—bipedalism first, brain second—was seemingly contradicted during the early 1900s with the discovery at Piltdown, England of a large, thin-boned, and rounded humanlike cranium, an apelike partial mandible preserving two molars, and an apelike lower canine. Under the presumption that these specimens were associated, the scenario of human evolution was turned around: Early human relatives became human first in their brains and then in the rest of the body (as inferred from the mandible and teeth). That is, the brain enlarged prior to the attainment of fully erect posture and bipedal locomotion. It was not until the 1950s, when the Piltdown fraud was exposed, that this alternative notion of human evolution—brain first, body second—was rejected. Before then, however, the discovery of the Taung child and Dart’s (1925) description and, more importantly, his interpretation of this specimen continued the intellectual trajectory Blumenbach had begun. But Dart conceived his scenario in the context of Darwin’s biogeographic premise of finding fossil evidence in Africa of intermediate forms that provided evidence of a morphological continuum between African apes and humans.

As Dart (1925 p 196) summarized his overall impression of the preserved craniodental features of the Taung specimen, this individual represented “an extinct race of apes intermediate between living anthropoids and man.” Dart depicted specific features—such as the configurations of the brow, nasal bones, zygomatic regions, orbits, and upper and lower jaws as well as the inferred skull shape—as being of “delicate and humanoid character” (Dart 1925 p 196). Perhaps most central to his speculations, however, were the size and potential details of the preserved partial endocast and also of the forward position of the foramen magnum (as indicated by bone adherent to the endocast). Dart (1925 p 197) assumed the latter was proof of this “humanoid’s” erect posture and then made the following extrapolations:

- The improved poise of the head, and the better posture of the whole body framework which accompanied this alteration in the angle at which its dominant member was supported, is of great significance. It means that a greater reliance was being placed by this group upon the feet as organs of progression, and that the hands were being freed from their more primitive function of accessory organs of locomotion. Bipedal animals, their hands were assuming a higher evolutionary role not only as delicate tactual, examining organs which were
adding copiously to the animal’s knowledge of its physical environment, but also as instruments of the growing intelligence in carrying out more elaborate, purposeful, and skilled movements, and as organs of offence and defence. The latter is rendered the more probable, in view, first of their failure to develop massive canines and hideous features, and secondly, of the fact that even living baboons and anthropoid apes can and do use sticks and stones as implements and as weapons of offence.

With regard to the Taung child’s brain (as represented by the endocast), Dart suggested that, since it was already as large as a chimpanzee’s and almost as large as a gorilla’s, it would have continued to enlarge, following a humanlike growth curve. In addition, as in humans but in contrast to apes, the Taung child’s brain was high and rounded, somewhat expanded in the temporal region, and bore an apparent lunate sulcus that was situated well posteriorly and inferiorly. On the basis of its seemingly more humanlike than apelike features, Dart inferred that this “humanoid” was humanlike in its faculties of “associative memory and intelligent activity.” The expanded cerebral cortex (as indicated by the supposed position of the lunate sulcus) also suggested to Dart that, in contrast to apes, the Taung “humanoid” had experienced increased sensory stimulation, both via vision (because of the forward position of the approximated orbits) and tactile sensation (because erect posture and bipedality supposedly freed the hands from involvement in locomotion). But the Taung child’s brain was not sufficiently enlarged in the temporal region for Dart (1925 p 198) to think that it had reached the “necessary milestone in the acquisition of articulate speech.”

As far as Dart was concerned, the Taung child, the name bearer of his new genus and species, *Australopithecus africanus*, displaced both Piltdown’s *Eoanthropus dawsoni* and Trinil’s *Pithecanthropus erectus* as viable “links” between humans and their apelike ancestors. Indeed, in spite of Dart’s statement about this juvenile extinct human relative’s intermediacy between humans and apes, his interpretation of it more closely reflected Blumenbach’s criteria for distinguishing *H. sapiens*. In 1925, then, while there were three somewhat but yet significantly different scenarios for the emergence of the earliest direct human ancestors, they all embraced the notion of an evolutionary continuum that extended from an apelike precursor, through a still unknown series of intermediates, to the most modern looking of living humans.

In terms of the focus of this chapter—defining Hominidae—further discoveries of potential extinct human relatives are less relevant than are attempts to integrate these fossils into a systematic framework that had originally been based on living *H. sapiens* alone. Thus, after a decade-and-a-half of successful fossil hunting in the limestone caves of South Africa and in caves and deposits in
Europe and Asia, a proliferation of genus and species names, and debates over the relationships to accepted hominids of the African taxa (referred to here as the australopiths), Clark (1940) was compelled to review the available evidence in order to determine the hominid status of any australopith.

In addition to echoing Huxley and Darwin’s assumptions of there having been a transformation from ape to human, Clark (1955) based his conclusions on what he later called the “total morphological pattern.” That is, he decided whether a particular fossil was hominid, not in terms of derived features it might share with humans but, rather, in terms of whether, overall, it resembled humans more than great apes. As will become obvious, this approach complicated matters further because the great apes were considered a group from which features for comparison could be extracted differentially from any ape, even if they characterized only one; Gregory (1922) had also used the device of “pick and choose” in arguing for an African ape–human relationship (Schwartz 2005). The irony of Clark’s using this phenetic approach is that, in 1955, he also made one of the clearest arguments for distinguishing between primitive and derived characters when making phylogenetic comparisons.

Although Clark (1940 p 317) concluded that australopiths were “more human than simian” especially in their teeth, his comparisons with apes were essentially restricted to the African taxa—which was a biased comparison he continued to emphasize throughout subsequent publications on fossil hominids. If he had included orangutans in his studies, he would, or at least should, have been struck by the similarities between this hominoid and australopiths in many details of facial and dental morphology (see Schwartz 2004a; also Schwartz and Tattersall 2005). This latter omission aside, Clark’s efforts to define Hominidae were as useless as his definition of the order Primates [i.e., primates are characterized by their lack, not sharing of any particular derived feature/s (Clark 1959)].

Mayr’s (1950) influential article on fossil hominids did not clarify the situation. Rather, on the grounds that all hominids were effectively adaptively similar because they were bipedal, Mayr collapsed all named taxa into one genus, Homo. Then after claiming that, since living humans are so diverse and occupy all available econiches the same must have been true for all hominids (thus precluding the opportunity and prerequisite for speciation), he subdivided his genus Homo into three time-sequential species: transvaalensis, erectus, and sapiens. Even though, in light of a much enlarged human fossil record, Mayr (1963) later “accepted” Australopithecus and made a nod toward Paranthropus as side branches that went extinct without issue, his concession did not elucidate how one decided in the first place if a specimen was a hominid, especially in the absence of postcranial remains. This is, indeed, a problem. For while it may be
true that some prominent scholars (e.g., Clark, Mayr, Washburn) decided “that the most important single factor in the evolutionary emergence of the Hominidae as a separate and independent line of development was related to the specialized functions of erect bipedal locomotion” (Clark 1964 p 14), the preponderance in the human fossil record of often fragmentary skulls and jaws and isolated teeth makes impossible identifying a specimen as being “hominid” on the basis of anatomical features believed to be reflective of bipedal locomotion. The exception, of course, is the region of the foramen magnum and occipital condyles.

2.3 Toward a definition of Hominidae

The task of defining Hominidae is twofold. First, there is a taxonomic decision. How widely does the classificatory net, Hominidae, extend? To chimpanzees? Chimpanzees and gorillas? All great apes? Although Clark (1955, 1964) wrote at a time when the great apes were believed to constitute a group or clade, and were thus classified together in the family Pongidae, his rationale for recognizing a group or clade Hominidae is, I believe, still viable and useful. Hominidae is a monophyletic group that subsumes extant humans and their fossil relatives to the exclusion of any living relative.

Some scholars who embrace molecular similarity between humans and chimpanzees as indicating evolutionary closeness then also group these hominoids in the same genus on the grounds that this endeavor is “cladistic” (Goldberg et al. 2003). Unfortunately, this taxonomic device hinders attempts to make sense of a clade (no matter what it is called) that subsumes humans and their closest relatives. For our closest relatives are not extant apes. They are as yet an unknown number of now extinct taxa. Furthermore, Goodman’s and like-minded molecular systematist’s claims to being “cladistic” in folding Pan and all australopiths into the genus Homo is relevant only in the context of translating a presumed scheme of phylogenetic relationship into a classification, as Hennig (1966) advocated. In truth, however, the asserted correlation between molecular identity and evolutionary propinquity is only an assumption of continual molecular change (Zuckerkandl and Pauling 1962; Caccone and Powell 1989). Furthermore, the aura of being “cladistic” is an artifact either of embracing this assumption and then stating that greater molecular similarity must therefore reflect a greater degree of shared derivedness or of rooting a tree in a taxon that is from the outset presumed to be primitive relative to those taxa of phylogenetic interest and then asserting that the similarities between the latter are synapomorphic (see discussion in Schwartz 1987, 2005a, 2005b). In reality, the only constant source of mutation is UV radiation, whose effects are not only very
infrequent (typically $10^{-8}$ to $10^{-9}$) but also random with regard to impacting somatic versus sex cells (the latter being the only cells relevant to questions of relatedness).

Even though the notion of DNA being the “blueprint of life” has long outlived its utility, there still seems to be a lingering intuitive “sense” that molecular similarity should be phylogenetically revealing and also somehow be reflective of morphological similarity. But in metazoans, RNA delineates different reading frames from the same stretch of DNA, and it and various classes of proteins orchestrate signal transduction pathways and maintain cellular and molecular homeostasis (Carroll 2003; Maresca and Schwartz 2006 Schwartz 2005a, b). Indeed, the very nature of a cell’s biology to edit out or at the very least to suppress potential “mutations” both calls into question the “molecular assumption” of continual and accumulative mutation (cf. Caccone and Powell 1989) and more properly directs attention to the molecular and cellular (including physical and epigenetic) interactions that ultimately give rise to a functioning organism (Schwartz 2005). Consequently, from a molecularly based developmental perspective, morphology is indeed a viable resource in determining phylogenetic relationships (Schwartz 1999b, 2005b).

The latter point is not trivial because, in various cases, including that of *H. sapiens*, most potential insight into the evolutionary history of a clade comes from the fossil record. And, of course, fossils represent the physical remains of organisms—the results of their developmental histories. It is, therefore, odd that paleoanthropologists who accept without question a close relationship between humans and chimpanzees—for which there is virtually no morphological support (cf. Schwartz 1988, 2005; Shoshani et al. 1996; *contra* Begun 1994; Begun et al. 1997), whose claim for demonstrating a human–chimpanzee on the basis of morphology is merely an extrapolation from an assumed relationship between “*Australopithecus*” (undefined) and *Pan* (cf. Conroy 1994)—and, consequently, reject morphology as being phylogenetically revealing can then turn to morphology in order to analyze fossils (Pilbeam 1986; Asfaw et al. 1999).

Nevertheless, this contradiction leads to the second issue in defining Hominidae: the criteria one uses. If we accept the “molecular assumption” and its consequences—that morphology, especially fossilizable morphology, is phylogenetically unrevealing (Collard and Wood 2000; Pilbeam 2000)—then we are left with no alternative other than to throw up our hands in despair and wonder in amazement at how similar in various distinctive ways the morphologies of what we have been identifying as fossil hominids truly are. Hominidae as a potential clade is reduced to the single genus and species Blumenbach sought to define: *H. sapiens*. Nothing that has been considered an extinct hominid can be accepted
as such. Of course, when put in these terms, this is a ludicrous suggestion. Nevertheless, it is the logical extension of the “molecular assumption.”

As Carroll (2003) pointed out while warning against a naive expectation of developmental and evolutionary “insight” when human and chimpanzee genomes are ultimately compared in detail, there is not a one-to-one correlation between DNA sequence and morphology. An organism arises as a result of the ways in which segments of DNA are differentially translated and introns differentially spliced and the resultant products and other molecules are recruited and deployed. Lovejoy et al. (1999) offered a compelling argument in this regard, one that is not unrelated to the task of defining Hominidae. As they suggested (p 13428) with regard to the emergence of the clearly derived human pelvic form, “if a particular PI [positional information] gradient were to span n cell diameters, and those cells defined the ultimate anteroposterior dimension of the presumptive ilium (superoinferior in the adult human), then a slight increase in the steepness of its slope would cause that signal to span fewer cells, ‘distorting’ the presumptive anlagen and substantially altering downstream adult morphology.” That is, although received wisdom may embrace a process of change involving myriad small steps, “the transformation of the common ancestral pelvis [in its entirety] into that of early hominids may have been as ‘simple’ as a slight modification of a gradient” (Lovejoy et al. 1999; comments added). Consequently, while we might not know every aspect of molecular signaling and cell interaction underlying the emergence of a complete organism, we can still hypothesize from a grounding in cell biology that similar morphologies result from similar molecular and cellular interactions (Schwartz 2005b). As such, we can continue to test our theories of homology as we have all along in systematics: with other theories of homology.

Lovejoy et al.’s (1999) developmentally informed suggestion also brings with it other considerations relevant to the task of defining Hominidae (or any clade). What we might correctly interpret as major morphological change is not necessarily due to incremental and cumulative transformation over a significant period of time but may result from an integrated cascade of change that produces large-scale novelty rapidly, not slowly. This, of course, is theoretically expected (Schwartz 1999a), especially given the cell’s inherent mechanisms for maintaining homeostasis and resisting change (Maresca and Schwartz 2006; Schwartz 2005b). That is, it is likely that extremes of cellular perturbation, and not a history of accumulated “genetic change” (whatever that actually means), may permit the expression of novel avenues of molecular communication and cellular interaction that are normally suppressed or eliminated (Maresca and Schwartz 2006).
Recognizing that the disruption of DNA and cellular homeostasis can lead to large-scale morphological change that is abruptly expressed [in multiple individuals (Schwartz 1999a)], and which, if not lethal, may persist (Maresca and Schwartz 2006; Schwartz 1999a), makes futile the search for evidence—fossil or otherwise—of a smooth, incrementally continuous transition from one character state to another or from one taxon to another (Schwartz 1999a). In addition, even a superficial appreciation of the interplay between molecular communication, cell interaction, and organismal development highlights the problems of infusing systematic methodology with expectations of “discovering” evidence of transformation series. I refer here particularly to the often-employed concepts of “incipient” and “vestiges of”: “incipient” implying that a feature is “on the way” to becoming something else or at least a different version of itself (e.g., an incipient supraorbital torus in the case of *Dryopithecus* being touted as an ancestor of African apes and hominids), and “vestiges of” reflecting the idea that, since morphological features evolved “in order to” perform a certain function, one might find hints of them in organisms that no longer engage or engaged in that activity, but whose ancestor did (e.g., a vestige of a facet on the distal ulna of some specimens attributed to *Australopithecus afarensis* that one finds occasionally in gorillas, wherein knuckle walking is taken as the ancestral condition for hominids).

With these caveats in mind, we might now turn to the matter of defining Hominidae but not from the perspective of looking for the “defining” moment in a transition from some apelike condition to something seemingly hominid (either by a subtle hint of a supposed hominid trait or traces of a presumably primitive and retained feature). Inevitably, we return to Blumenbach’s list of criteria, to which other features have subsequently been offered up as delineating a clade that I shall refer to as Hominidae: humans and their extinct relatives.

2.4 Defining characters of Hominidae?

2.4.1 Traditionally accepted features of “erect posture”

As reviewed above, Blumenbach’s emphasis on “erect posture” and “two handedness”—or, as Clark (1964 p 14) put it, on “specialized functions of erect bipedal locomotion”—has loomed large in considerations of the constitution of our own clade. Pilbeam (1972 p 62), for example, summarized some of the “adaptations” that are apparently associated with these “specialized functions of erect bipedal locomotion”: a vertebral column with a distinct lumbar curve that is set at a sharp angle relative to the sacrum; a “carrying angle,” wherein the lateral femoral condyle is larger and more weight bearing than the medial one, and the shaft of
the femur angles up and laterally away from the knee joint; a nongrasping foot with short toes and nondivergent hallux through which weight is transmitted during locomotion; and metacarpals in which the heads contact the substrate while the distal ends are elevated to form a springlike, transverse arch. We might also include Blumenbach’s description of the pelvic region as being bowl shaped (i.e., broad and shallow) and having a short, potentially laterally flaring, posteriorly expanded, anteriorly truncated ilium, a somewhat forwardly oriented acetabulum (which is also reflected in the orientation of the proximal femur relative to the shaft), a defined greater sciatic notch, and a broad, short sacrum, wherein the alae are not remarkably small relative to the size of the lumbar facet (Clark 1964; Schultz 1968). Clearly, these features distinguish living *H. sapiens* from other extant primates. However, the degree to which these characteristics are expressed in what have been identified as fossil hominids, and whether the appropriate postcranial remains are known, leaves much open for questioning.

For instance, among fossil specimens attributed to “anatomically modern” *H. sapiens* that probably represent this taxon (Schwartz and Tattersall 2000a, b, 2003) only Qafzeh 9 is known from a fairly complete, albeit, extremely crushed postcranium. Inasmuch as distortion of the skull and mandible compromises the definitive identification of a bipartite brow with a “glabellar butterfly” and a “true” chin with an inverted “T” configuration and thickened inferior symphyseal margin (Schwartz and Tattersall 2000a, b, 2003), postcranially, the pelvic region appears to have the details of *H. sapiens* (personal observation). Other specimens attributed to “anatomically modern” *H. sapiens*—especially Qafzeh 6 and all from Skhul—with sufficiently preserved crania and/or mandibles do not present a bipartite brow or an inverted mandibular symphyseal “T” (Schwartz and Tattersall 2000a, b, 2003). Known Skhul postcranial remains are incompletely representative and so crushed and poorly reconstructed that one can merely get a sense of their conforming to the above-mentioned pelvic configurations (personal observations).

Collectively, while Neanderthal postcrania (especially those associated with bipedal locomotion) differ from those of *H. sapiens* in details of size, shape, and/or morphology [e.g., more posteriorly expanded ilia, superoinferiorly tall and anteroposteriorly compressed pubic symphyseal region, relatively very elongate pubic ramus (and thus very wide subpubic angle), smoothly “hook-shaped” greater sciatic notch, smaller and differently oriented iliac auricular region, relatively large proximal and distal femoral ends, very large acetabulum, truncated calcaneus (Trinkaus and Howells 1979; Trinkaus 1983; Rak 1990; G. Sawyer reconstruction, personal observations)], they can still be accommodated by the general configurations summarized above. In addition, Neanderthals had a “carrying angle” and a lumbar curve.
With regard to Middle Pleistocene fossils, aside from the unique collection derived from Atapuerca’s Sima de los Huesos (Arsuaga et al. 1997), there are few instances where postcrania are associated with craniodental remains, Arago 44 (Day 1982; Sigmon 1982, personal observation) and the Jinniushan specimen (Rosenberg and Lu 1997) being obvious exceptions. With regard to the task of defining Hominidae, information on the pelvic region is the most readily available. Although differing from *H. sapiens* especially in some details of the ilium (e.g., flare, anterior superior iliac spine), Pilbeam’s criteria are still applicable to these fossils, as is also a femoral carrying angle in Sima specimens (Day 1986, personal observation; Rosenberg and Lu 1997).

Among earlier Pleistocene hominids, the Nariokotome (KNM-WT 15000) femora, vertebrae, and pelvic region, the Olduvai OH 8 foot bones, the Trinil and various other femora (especially Koobi Fora KNM-ER 1481), and the OH 28 and other os coxae (KNM-ER 1808 and 3228) are most relevant. Rose (1984), for example, compared the ER 1808 and 3228 favorably with those from Arago and Olduvai and also noted that they differ from *H. sapiens* and *H. neanderthalensis* in being relatively larger in the anterior region of the ilium. Although this region is not well preserved in either os coxa of KNM-WT 15000, it seems to compare well Arago 44, OH 28, and ER 1808 and 3228 (Day, 1986; Walker and Leakey 1993). An interesting but not especially emphasized aspects of these ilia is that, when the acetabula are oriented laterally, the “inner” surface of the blade faces noticeably forward (as in Sts 14), rather than medially as in *H. sapiens* and *H. neanderthalensis* (and also SK 50). The lumbar region of WT 15000 is somewhat curved, and there is a distinct angle at L5-S1 (cf. reconstruction in Walker and Leakey 1993). With regard to femora, WT 15000, ER 1481, and the complete Trinil specimen present a carrying angle, and in all the head is large. Interestingly, however, while the femoral neck is short and rounded in cross section in ER 1481 and Trinil, it is relatively long and somewhat flattened anteroposteriorly in both WT 15000 femora. The OH 8 foot bones, while reconstructed by Day and Napier (1964) with an arch and a first digit that was aligned with the other digits, may instead have had a semiopposable hallux (Clarke and Tobias 1995) (thereby suggesting that the proximal ends of the metatarsals were not elevated into an arch). If the OH 8 foot bones are those of a “hominid” (they were not associated with other skeletal or dental elements), then the pedal features in Pilbeam’s list do not define this potential clade.

Among Plio-Pleistocene specimens, the lumbar region and apparently also the lumbosacral articulation of Sts 14 essentially conform to Pilbeam’s criteria (Robinson 1972). In addition, the ilia of Sts 14 are both low and squat superoinferiorly, roundedly expanded posteriorly (and thus appear long anteroposteriorly) and have a well-defined greater sciatic notch. The preserved left os coxa
from the small Hadar AL 288-1 and the Makapansgat juvenile ilia (MLD 7 and 25) are similar to Sts 14 in these features, as apparently was also the right partial os coxa from Swartkrans (SK 50), which lacks the regions of the iliac crest and pubis (Robinson, 1972; Johanson et al. 1982; Lovejoy et al. 1982). With regard to adult ilia, at least Sts 14 and AL 288-1 and possibly also SK 50 differ from specimens discussed above in being oriented more laterally than vertically (thus exposing upward more of the internal surface of the ilium) and in having less blatantly “S”-shaped iliac crests; in addition, these specimens are only somewhat thickened in the region of an iliac (crest) tubercle and bear poorly developed iliac pillars (cf. Robinson 1972; Johanson et al. 1982; Day 1986). The preserved pubic rami of Sts 14 and AL 288-1 are reminiscent of those of Neanderthals and Jinniushan in being relatively long, but the symphyseal regions of the former two specimens are not also superoinferiorly tall and anteroposteriorly compressed (cf. Robinson 1972; Johanson et al. 1982; Day 1986; Rosenberg 1998). A superior view of the articulated Sts 14 pelvis illustrates (in contrast to \textit{H. sapiens}) the relation of the elongate pubic rami to the relatively wide pelvic canal and the relatively posterior positioning of the outwardly flared ilia (cf. Robinson 1972). Furthermore, while in \textit{H. sapiens} (and other Pleistocene specimens surveyed above) the curve of the iliac crest positions the anterior superior iliac spine just lateral to the parasagittal plane that intersects the posterior superior iliac spine, in Sts 14, the anterior portion of the ilium would have lain well lateral to the plane of the sacroiliac articulation (Robinson 1972). Even though only the left os coxa of AL 288-1 is known, it was likely similar to Sts 14 (cf. Johanson et al. 1982). In spite of the fact that the iliac crest of SK 80 is similar to Sts 14 in not being strongly “S-shaped,” when their ilial blades are oriented in the same plane, the acetabulum of SK 80 is similar to \textit{H. sapiens} in facing laterally and slightly downward (Robinson 1972). In contrast, the fairly vertically aligned acetabulum of Sts 14 faces forward. When compared in the anatomical position, the Sts 14 ilium again assumes a more outwardly oblique orientation while SK 50 is more anteroposteriorly oriented, as in \textit{H. sapiens} and various other specimens of “\textit{Homo}.”

The anterior part of australopith ilia present a dichotomy of morphology. In the better preserved left os coxa of Sts 14, this region appears to be roundedly expanded anteriorly (Robinson 1972) as in StW 431 (Toussaint et al. 2003) and AL 288-1 (Johanson et al. 1982). Thus, a definitive anterior superior iliac spine cannot be identified. But in SK 50, even though its iliac crest is damaged along much of its length, what is preserved continues forward to become a well-defined, beak-shaped anterior superior iliac spine that projects markedly anterior to a bluntly thickened, almost knoblike anterior inferior iliac spine (cf. Day 1986; Robinson 1972). The smaller Makapansgat juvenile ilia (MLD 7 and 25) are similar to SK 50 (but not to
Sts 14 and AL 288-1) in having a projecting, beaklike anterior superior iliac spine (cf. Dart 1949, 1957; Robinson 1972). Although differing from SK 50 and MLD 7 and 25 in the region of the anterior superior iliac spine, Sts 14 and AL 288-1 are similar to them in developing a knoblike anterior inferior iliac spine that lies noticeably superior to and back over the superior margin of the acetabulum, as in other potential hominids surveyed above.

As is well known, australopiths have traditionally been interpreted as being postcranially intermediate between knuckle-walking great apes and “Homo”—as noted, for instance, in their developing a humanlike posterior iliac expansion while supposedly retaining an apelike anterior iliac distension. Given the above comparisons, this scenario seems inappropriate. Not all australopiths are characterized by the same configuration of the region of the anterior superior iliac spine. Only SK 50 and MLD 7 and 25 compare at all favorably with great apes in having a beaklike anterior superior iliac spine that continues forward the trajectory of the iliac crest. The rounded anterior expansion of this region in Sts 14 and AL 288-1, while absolutely and relatively large compared to *H. sapiens* and other possible fossil hominids, is, nevertheless, derived in its own right (Figure 2.1).

In contrast to hominoids in general, the acetabula of Sts 14, AL 288-1, and SK 50 are relatively and absolutely small (Robinson 1972; Day 1982; Johanson et al. 1982). Also in contrast to hominids in general, the head of the associated femur of AL 288-1 is relatively small and does not extend much circumferentially beyond the dimensions of the neck (especially in lateral view), and the femoral neck is

---

**Figure 2.1**

Os coxae of Sterkfontein Sts 14 (cast) and *H. sapiens*. Although both have a knoblike anterior inferior iliac spine, its position differs, as do the orientations of the ilia and acetabula (which is relatively smaller in the extinct hominid). See text for further discussion; not to scale. (© J. H. Schwartz)
long and somewhat flattened or compressed anteroposteriorly (Lovejoy et al. 1982). Other preserved proximal femora (e.g., AL 128-1, 333-3 and 333-95, SK 82 and 97, BOU-VP-12/1, and KNM-ER 738 and 1503 and apparently also 1475) are essentially similar to AL 288-1 (cf. Napier 1964; Day 1986 (Figure 2.2)). The Maka, Bouri (BOU-VP-12/1), and Sts 14 proximal femoral fragments lack the

head, but in preserved parts their necks are similar to the specimens just mentioned (cf. Robinson 1972; White et al. 1993; de Heinzelin et al. 1999). Since the broader comparison indicates that a long, compressed femoral neck is derived among hominoids, it may be phylogenetically significant that the femoral neck of the Nariokotome specimen (KNM-WT 15000) is elongate (Walker and Leakey 1993), as is also the neck of the proximal femur (BAR 1002'00) from the Late Miocene site Lukeino, Kenya that Pickford et al. (2002) referred to Orrorin tugenensis [the type specimen of which is an unassociated jaw with a few teeth (Senut et al. 2001)] (Figure 2.3).

Specimens with a “carrying angle” (characterized by the development of distal femoral condyles of different size and orientations and a laterally slanted femoral shaft) are known, for example, from Sterkfontein (Sts 34), Kromdraai (TM 1513), Hadar (AL 288-1), and Koobi Fora (e.g., KNM-ER 1475). The lateral slant of the femoral shaft is most pronounced in the former three specimens, but, in all, it is much more marked than in specimens or taxa discussed earlier.
To review, there are some features of the os coxa and femur that, in contrast to at least other catarrhines, appear to distinguish a subgroup of hominoids that we might refer to as “hominid”: i.e., a long pubic ramus, a superoinferiorly short ilium that is roundedly expanded posteriorly, some thickening in the region of an iliac (crest) tubercle, a well-developed and knoblike anterior inferior iliac spine that lies noticeably superior to and somewhat back over the superior acetabular rim, a deep greater sciatic notch, a defined linea aspera, femoral condyles that are of unequal size and different orientations with an outwardly slanted femoral shaft (producing a “carrying angle”), and a concave lateral tibial facet for the femur that is at the same level as the (also but primitively concave) medial facet, with the two facets being separated by well-developed tubercles.

Within this potential hominid clade, a subclade is distinguished by having an elongate and somewhat compressed femoral neck. The larger femoral head of specimens, such as WT 15000, ER 1547, and BAR 1002'00, is interesting in light of

---

1 Pickford et al. (2002) also argued that a greatly elongate femoral neck was a derived condition among hominoids but concluded that this was a shared derived state for Hominidae, not, as its more restricted distribution would suggest, a shared derived state of a less inclusive clade.
the possibility of their belonging to a clade united by femoral neck elongation: either these specimens and others like them are derived within the latter clade or, if their relationships lie with specimens attributed to “Homo,” femoral neck elongation is independently derived for them.

Before continuing, two features should be considered. One is the configuration of the distal femur when viewed from below (Figure 2.4). As initially illustrated by Robinson (1972), while TM 1513 and Sts 34 do not differ greatly from great apes in disparity and/or orientation of medial versus lateral condyles, the fossils are more strikingly different in being deeper anteroposteriorly and more trapezoidal (mediolaterally narrower anteriorly) rather than rectangular in outline. In further contrast to great apes, the patellar surfaces of TM 1513 and Sts 34 are somewhat more concave and slightly asymmetrical (i.e., in the anatomical position, the lateral margin of the patellar sulcus is more anteriorly elevated than the medial margin). The latter disparity is more obvious in TM 1513 than it is in Sts 34. The AL 288-1 distal femur is more similar to TM 1513 than to Sts 34 (cf. Lague 2002). In H. sapiens and H. neanderthalensis distal femora, for example, there is greater disparity between size and orientation of the medial versus lateral condyle, the bone is mediolaterally narrower and anteroposteriorly deeper, the patellar sulcus is deeper (and thus more roundedly “V” than broadly “U” shaped), and the lateral margin of this surface is much more noticeably protrusive anteriorly. As such, perhaps the features of the distal femur as particularly represented in Sts 34 had characterized the last common ancestor of a clade Hominidae, while the configuration of H. sapiens distinguished the ancestor of a subclade. Interestingly, the outline and specific configurations of the condyles and patellar sulcus as well as the margins of the preserved distal end of the left femur of WT 15000 more closely resemble australopiths (especially Sts 34) than H. sapiens.

Figure 2.4
Distal view of right femora of (from left to right) P. troglodytes, Kromdraai TM 1513 (cast), and H. sapiens. Note differences in relative sizes of medial and lateral condyles and development of differently shaped patellar surfaces in the extinct versus extant hominid. See text for further discussion; not to scale. (© J. H. Schwartz)
The other feature to consider is the distance between the ischial tuberosity and the inferior acetabular lip, which is quite pronounced in great apes and apparently catarrhines in general (cf. Robinson 1972, personal observations; Aiello and Dean 1999). The separation between these two anatomical regions is marked in SK 50, shorter in Sts 34, and minimal in *H. sapiens*, in which a deep groove intervenes between the two structures. Interestingly, in WT 15000, the distance between these structures is reminiscent of Sts 34. Perhaps further study of this region will prove enlightening, if not in further defining features of Hominidae, perhaps in delineating a subclade within Hominidae.

### 2.4.2 More recently suggested postcranial features of “erect posture”

With the discovery at the Kenyan sites of Kanapoi and Allia Bay of specimens attributed to the species *Australopithecus anamensis*, other postcranial features were highlighted as distinguishing hominids from apes and, by extension, from other catarrhines (Leakey et al. 1995). Of particular note for this chapter is the configuration of the proximal tibia. In apes, the tibial facet for the medial femoral condyle is somewhat convex, while in specimens attributed to “accepted” australopith taxa and to species of *Homo* this facet is somewhat concave. In addition, the medial and lateral proximal tibial facets are separated by well-defined tubercles. One might also add to the tibial criteria for distinguishing between apes and potential hominids the fact that the proximal tibial facets are level with one another in “australopiths” and “Homo,” whereas in apes, the lateral facet lies below the level of the medial one (Aiello and Dean 1999; personal observations) (Figure 2.5).

![Figure 2.5](image)

**Figure 2.5**
Proximal tibiae, posterior view (from left to right) of *P. troglodytes*, Hadar AL 288-1A (cast), Kanapoi KNM-ER 29285 (cast), and *H. sapiens*. In the ape, note that the lateral condylar facet is convex anteroposteriorly and lies below the level of the shallowly concave medial condylar facet. In the hominids, the lateral condylar facet is concave and situated almost level with the medial condylar facet. See text for further discussion; not to scale. (© J. H. Schwartz)
In their description of the Lukeino proximal femur (BAR 1002'00), Pickford et al. (2002) commented on how slight the linea aspera is compared to specimens of australopiths and Homo (because the spiral line of BAR 1002'00 does not meet the linea aspera to form a structure of high relief). I agree (personal observation). They (p 202) then present “a precursor of the linea aspera” as a feature of bipedalism they would use to identify a femur as being hominid. Since, in BAR 1002'00, this faint muscle scar descends from the region of a gluteal tuberosity that lies below the greater trochanter, one can indeed distinguish this specimen, “australopiths,” and “Homo” from other hominoids. Apes do not present a muscle scar descending from the region of the gluteal tuberosity—which suggests that perhaps a “hint of a linea aspera” might define Hominidae.

Pickford et al. (2002) also mentioned that the lesser trochanter of BAR 1002'00 is medially projecting and, as in humans, gorillas, and Pongo, well separated from the femoral neck. Yet even among the illustrations in their article as well as those in Robinson (1972), Day (1986), and Walker and Leakey (1993), for instance, it appears that while a notable separation of lesser trochanter and femoral neck may characterize specimens such as SK 82 and 97, Sts 14, AL 288-1, KNM-WT 15000, and KNM-ER 738, 1503, and 1547, this configuration does not obtain to H. sapiens (or, for example, to Pan either). Perhaps this is due to differences in femoral neck length—which might imply that separation of the lesser trochanter from the femoral neck is phylogenetically significant within a clade Hominidae. The medial orientation of the lesser trochanter of BAR 1002'00, although similar to H. sapiens, is also characteristic of primates in general (cf. Swindler and Wood 1973; Aiello and Dean 1999; personal observation). In contrast, in specimens such as AL 128-1, 288-1, AL 333-95, AL 333-3, Maka VP 1/1, SK 82 and 97, Sts 14, WT 15000, and ER 738 and 1547, the lesser trochanter is posteriorly facing. Although not a defining character of Hominidae, perhaps the latter is a derived feature of a hominid subclade.

Among other features, Pickford et al. (2002; also Galik et al. 2004) suggested unite Orrorin via BAR 1002'00 with australopiths and at least H. sapiens, differential distribution of cortical bone of the femoral neck—thinner superiorly, thicker inferiorly—emerges as a potential defining characteristic of Hominidae (cf. Ohman et al. 1997).

2.4.3 Nonpostcranial features of “erect posture”

Although there are, no doubt, other features of the postcranium that might delineate a clade Hominidae, it might be prudent to turn now to another skeletal region from which “erect posture” or “bipedal locomotion” has been inferred: the base of the skull. For, ever since Dart’s (1925) discussion of the Taung child, a
forward position of the foramen magnum and its attendant occipital condyles has been central to the interpretation of australopiths and other early, potential hominids as bipedal hominoids. With Arditiphecus ramidus (White et al. 1994) and Sahelanthropus tchadensis (Brunet et al. 2002) being promoted as hominids, attention to the basicranium has never been more critical.

The interpretation of Arditiphecus (White et al. 1994) and Sahelanthropus (Brunet et al. 2002) as “hominid” was in part based on White et al.’s inference of bipedalism from the intersection of the bicarotid (foramen) cord and basion (on the anteriormost margin of the foramen magnum). The published photograph of Sahelanthropus’s crushed and distorted basicranium reveals, however, that what appears to be the anterior margin of the foramen magnum lies posterior to the bicarotid cord. In the undistorted state, the preserved left petrosal, and thus the bicarotid cord, would clearly have been more anteriorly placed.

Although not demonstrated, White et al.’s assertion of an association between basion and the bicarotid cord, an anteriorly placed foramen magnum, and erect posture and bipedalism seems intuitively likely. However, comparison of other potential fossil hominids, extant large-bodied hominoids, and various extant New and Old World monkeys, reveals a much more complex picture (Schwartz 2004a; Schwartz and Tattersall 2005; see also Wolpoff et al. 2002). In juvenile anthropoids, including humans, basion, the bicarotid cord, and the biporionic cord are essentially in alignment. This relationship is retained in the adults of some taxa, but, in others, the positions of basion and/or the bicarotid cord may change with growth relative to the biporionic cord. Consequently, it is not the bicarotid but the biporionic cord that appears to reflect the position of the foramen magnum. The alignment of basion and the two cords in the adult is, therefore, a neotenic feature, which, while not defining a clade Hominidae (Schwartz 2004a; Schwartz and Tattersall 2005), may be relevant to the delineation of relationships within it.

2.4.4 Proposed craniodental features of being hominid

Based on isolated teeth attributed to Arditiphecus, White et al. (1994) suggested that, in side profile, a hominid’s permanent upper canine has subequally long and quite divergent mesial and distal edges terminating in “shoulderlike” basal swellings that create the impression of a superoinferiorly short crown. This does not, however, describe the permanent upper canine (C1) preserved in the skull of Sahelanthropus (Brunet et al. 2002) or of the majority of C’s of traditionally accepted Plio-Pleistocene and later hominids, including H. sapiens (cf. Schwartz and Tattersall 2005). Rather, White et al.’s description better captures the morphology of the C1’s of adult female orangutans and the deciduous upper canines of juvenile orangutans and chimpanzees (cf. Swarts 1988).
White et al. (1994) then described *Ardipithecus* C1 as “incisiform,” which derived from the belief, also shared by Brunet et al. (2002), that a C1-C1-P1 honing complex had been lost in hominids via a decrease in size and trenchantness of these teeth (particularly the canines) and closure of associated diastemata. According to the scenario, as the canines became less caniniform, they became increasingly associated functionally with the incisors, ultimately assuming the morphology and function of incisors. Yet, both the right C1 allocated to *Ardipithecus* and the C1 in the skull of *Sahelanthropus*, although differing in the triangularity of their buccal outlines, are apically pointed. Indeed, *Sahelanthropus*’ trenchant C1 would likely have occluded with a much more “caniniform”-looking C1 than the very unprimatelike isolated tooth thusly identified (Schwartz 2004a). Although it appears from teeth in situ that *Sahelanthropus* lacked upper diastemata, their absence is inferred for *Ardipithecus* from isolated teeth. Nevertheless, if *Ardipithecus* did not have diastemata, it and *Sahelanthropus* would be more derived than geologically younger (and supposedly descendent) specimens with diastemata [e.g., Hadar AL 200-1a and Sterkfontein StW 252 (Schwartz and Tattersall 2005); the maxilla from Bouri Hata allocated to *Australopithecus garhi* (photographs in Asfaw et al. 1999)]. Consequently, features associated with “reduction of a canine-premolar honing complex” do not define a clade Hominidae, but they might be delineate subclades within it.

Another approach to defining Hominidae is predicated on an a priori assumption of a human– chimpanzee relationship, which immediately constrains the hominin “outgroup” to a single large-bodied hominoid, i.e., *Pan*. Thus, although White et al. (1994 p 306) describe the dm1 of *Ardipithecus* as “apelike,” the only ape in their comparison was *Pan*. Perhaps not surprisingly, the dm1s of *Hylobates* and *Gorilla* are very similar to *Pan*. But *Pongo*’s differently configured dm1 is actually similar to its counterpart in traditionally accepted hominids, the major difference lying in the ape having more talonid cusp compression (cf. Schwartz 2004a; Swarts 1988). Otherwise, the dm1s of *Pongo*, “australopiths” and “*Homo*” conform to White et al.’s (1994 p 307) depiction of “apparently derived hominin features”: i.e., “buccolingual crown expansion, mesiolyingually prominent metaconid, well-defined anterior fovea, and large talonid with well differentiated cusps” (White et al. 1994 p 307). In fact, the dm1s of *Pongo* and hominids2 are similarly derived compared to other extant hominoids in that the protoconid is more mesially situated; the less vertical and lingually facing anterior

---

2 The comparison includes juvenile “*Homo*,” including Melka Konturé MK81 GAR IV (2) (see Schwartz 1995; Schwartz and Tattersall 2002, 2003) and “australopiths” [e.g. Hadar AL 333-43b, Koobi Fora KMN-ER 820, 1477, and 1507, Laetoli LH2 and 3q, Omo 227, Taung, Swartkrans SK 47, Kromdraai TM1536, TM1601a, TM1604, KB5503, and KB5223, and Sterkfontein Sts24a (see Schwartz and Tattersall, in press)].
fovea (=trigonid basin) is noticeably smaller than the talonid basin and is enclosed by a distinct paracristid that courses somewhat mesially and then turns toward the lingual side; and the more horizontally oriented talonid basin is enclosed by a distinct hypocristid (personal observation; cf. Swarts 1988).

Only if *Pongo* (and other members of its clade for which dm1s are currently unknown, but for whom this description would be predicted as applicable) is not the sister taxon of a potential hominid clade would these dental features be relevant to defining the latter clade alone.

### 2.5 Additional defining characters of Hominidae?

The latter point is not trivial. For although some clearly derived features might in isolation delineate a clade Hominidae, they are also shared with *Pongo* and its potential extinct relatives. Consequently, in order to claim them as hominid apomorphies alone, and their presence in an orangutan clade as autapomorphic for it and nonsynapomorphic with hominids (i.e., homoplastically), one must first embrace another theory of extant large-bodied hominoid relationship and then “explain away” the phylogenetic significance of the similarities between humans and orangutans. It might therefore be useful to review briefly some of these features.

#### 2.5.1 *Pongo*-like dental and palatal features of potential hominids

An interesting case of what could be called “interpreter’s bias” lies in analyses of molar enamel thickness. For example, although Martin (1985) was cited as demonstrating that the last common ancestor of large-bodied hominoids had thick molar enamel (which was retained in orangutan and hominid clades but secondarily reduced in African apes), he in fact interpreted the enamel thickness data in the context of an (orangutan (human–African ape)) theory of relatedness (Schwartz 1987). Otherwise he should, or at least could, have concluded that thick molar enamel united human and orangutan clades. Pilbeam (1986; also Kelley and Pilbeam 1986) also interpreted molar enamel thickness in the context of an (orangutan (human–African ape)) theory of relatedness but presented two scenarios: Martin’s and one in which the last common ancestors of separate human and orangutan clades independently developed thick molar enamel. In the latter case, thick molar enamel would be a defining feature of Hominidae.
Most recently, Schwartz (2000) found that, while humans and orangutans both have thick molar enamel, humans have thicker enamel in some areas of their occlusal surfaces. He suggested that thick molar enamel had evolved independently in humans and orangutans (and, by implication, in the last common ancestors of separate human and orangutan clades) because of the demands of different diets. A more straightforward interpretation of Schwartz’s data is that human and orangutan clades shared a common ancestor that possessed thick molar enamel and that, within this clade, humans (and perhaps other hominids) are more derived. Only by accepting thick molar enamel as synapomorphic of traditionally accepted hominids can one embrace thin-enameled *Ardipithecus* as a sister of this clade (assuming that this relationship is based on synapomorphy). Of course, Brunet et al.’s (2002) suggestion that the thicker enameled *Sahelanthropus* is ancestral to *Ardipithecus* underscores the need for systematic rigor in paleoanthropology.

If thick molar enamel does not contribute to defining Hominidae, does dental morphology? Some aspects of it might, but, at present, delineating any would be a Herculean task. This difficulty derives from the longstanding notion of the Miocene being a time of an “ape radiation,” after which “hominids” emerged. Guided by this scenario, paleoanthropologists typically identified post-5.5 Ma specimens with thick-enameled cheek teeth as “hominid” and pre-5.5 Ma specimens as orangutan relatives. Nevertheless, study of fossils identified as hominid reveals that *Pongo*-like teeth have often been referred to species of “*Homo*” and “*Australopithecus*” (cf. Schwartz 2004b; Schwartz and Tattersall 1996; cf. Schwartz et al. 1995; Schwartz and Tattersall 1996, 2003; Schwartz 2004b; 2005a). Teeth allocated to *Orrorin tugenensis* (Senut et al. 2001), if truly thick-enameled, reinforce the question of how many Miocene “apes” are actually hominids and how many Plio-Pleistocene “hominids” are not. In addition, since “hominid” teeth are often worn flat, little attention has been paid to details of occlusal morphology. Thus, although most of the human fossil record has now been scrutinized at the level of morphs (Schwartz and Tattersall 2002, 2003, in press), much work remains in testing these hypotheses before assigning specimens (other than type specimens) to specific genera and species.

Were it not for their presence in *Pongo*, *Sivapithecus*, and *Ankarapithecus*, for instance, other features that could “define” Hominidae are the development of a single incisive foramen (Schwartz 1983, 1997) and a posteriorly thickened palate (as seen in midline cross section) (Schwartz 2004b). With regard to the latter feature, although it appears that the Hadar broken palate AL 200-1a thinned posteriorly, if true, this would be the outlier among “accepted” hominids (Schwartz 2004b; cf. Schwartz and Tattersall 2002, 2003, 2005). As with thick molar enamel and the morphological details of dm1, the simplest phylogenetic
interpretation of these features is that they characterized the last common ancestor of human and orangutan clades.

**2.5.2 Pongo-like features of “australopiths”: implications for defining Hominidae**

Another stumbling block to defining Hominidae is the degree to which “australopiths” and specimens traditionally allocated to “*Homo*” differ from one another, especially in craniofacial morphology: e.g., in orbital and nasal aperture outline as well as the configuration of the supra- and infraorbital regions; elevation of the nasoalveolar clivus above the floor of the nasal cavity; height and orientation of the infranasal region; orientation, flatness, and height of the zygomatic region; the development of a “snout” with/without facial pillars (vertical or medially inclined) and a canine fossa; and in the development and configuration of a mastoid process. Thus, while one might unite all of these potential hominids via various postcranial features, craniofacial morphology seems only to delineate possible hominid subclades. This may in part explain why one approach to linking humans with *Pan* is to try first to link the latter with australopiths and then to assert that, since australopiths and *Homo* form a clade, humans and chimpanzees are closely related (Begun 1994, Begun et al. 1997). In the latter case, the primary feature of supposed synapomorphy between (an undefined) *Australopithecus* and *Pan*—barlike supraorbital torus with sulcus behind—describes no “australopith” and only some specimens of “*Homo*” (Kimbel 1986; Clarke 1987; Schwartz 1997; Schwartz and Tattersall 2002, 2003, 2005). In addition, the markedly inferosuperiorly tall supraorbital region of *Sahelanthropus* is not barlike, but it is so unusually tall that it must represent a derived, not primitive, configuration—which, together with its distinct dental morphology and not very thick molar enamel, makes determining its relationship to other hominoids difficult (Schwartz 2004a).

While seeking connections between humans and African apes (especially *Pan*) has historical precedent (Clark 1940; Johanson and White 1979), the most favorable comparisons are actually between *Pongo* (and its fossil relatives) and australopiths (Schwartz 2004b, 2005a): e.g., rimlike superior orbital margins with a rather smooth transition into the frontal plane; often ovoid orbits; and forwardly facing, tall, and often vertical zygomatic regions (a further derived configuration, e.g., as in KNM-WT 17000, is a posteriorly tilted zygoma). Many australopiths are also similar to *Pongo* in having distinct facial pillars that emerge just above the canines and, together with a variably developed canine fossa, delineate a “snout.” Interestingly, the Taung specimen has the long, slitlike
single incisive foramen seen in *Pongo* and also preserved in *Sivapithecus* and *Ankarapithecus* (Schwartz 2005). While I am not, however, advocating a special relationship between a *Pongo* clade and some or all australopiths, attention to these apparent synapomorphies is more relevant to the task of defining Hominidae, its subclades, and sister taxa than is currently appreciated.

### 2.6 Final comments

When comparing extant taxa, defining Hominidae is a simple matter. Through a curious historical twist, *H. sapiens* is the only survivor of a clade whose dimensions we still do not fully know. Take, for instance, the recent discovery of a possible hominin from the Late Pleistocene of Flores, Indonesia (Brown et al. 2004). Its combination of morphologies would be unexpected at any time period.

It has a moderately globular cranium (as in some hominids), somewhat thickened and anteriorly (but not superiorly) protruding but rimlike supraorbital margins with no sulcus behind (australopiths in part and orangutans and their relatives), tall, ovoid orbits (*Pongo, Sivapithecus*, some “australopiths”), flat nasal bones (most hominoids, including some “*Homo*”), forwardly facing and vertical yet superoinferiorly short zygomas (australopiths and orangutans and their relatives), well-developed mastoid processes (some “australopiths” and some “*Homo*”), a thick frontal with thick diploe (autapomorphic or pathological), no frontal sinuses (most primates, including bonobos), an anterior cranial fossa that does not extend fully over the orbital cones (most primates, including some “*Homo*”), a clivus that slopes gently away from the dorsum sellae (most primates, including some “*Homo*”), basion, the bicarotid and biperionic cords in alignment (juvenile anthropoids, some adults), a broad incisive foramen that proceeds anteriorly as a expanding groove (a few “australopiths”), marked separation of the nasoalveolar clivus and an anteriorly thin palate (African apes, some Miocene hominoids, some “australopiths”), a smoothly but narrowly curved mandibular symphyseal region (many anthropoids), a long retromolar space (various “*Homo*”), a broadly and smoothly rounded but somewhat truncated gonial angle (some “*Homo*”), a very anteroposteriorly long sigmoid notch (some “*Homo*”), a sigmoid notch crest that is deepest near the coronoid process (autapomorphic), very large cheek teeth and apparently relatively small anterior teeth (some “australopiths”), unusually mesiodistally short upper and lower molars with large mesial and truncated distal cusps (autapomorphic), a relatively long ilium (SK 50) with a beaklike anterior superior iliac spine (great apes, SK 50, MLD 7 and 25), a poorly defined iliac pillar (australopiths), a knoblike anterior
inferior iliac spine that lies noticeably superior to above and somewhat back over
the supraacetabular rim (hominids as discussed here), an arcuate line that
descends well before reaching the region of the acetabulum (some hominids),
no ischial spine (most primates), posterior iliac expansion that defines a greater
sciatic notch (hominids as used here), a “V”-shaped greater sciatic notch (most
hominids, but not Neanderthals), a large femoral head (most primates), a long
and anteroposteriorly compressed femoral neck (Orrorin “australopiths” includ-
ing BOU-VP-12/1, WT 15000), a well-defined intertrochanteric crest (most
primates, but not WT 15000, Orrorin, and most if not all “australopiths”), a
medially facing lesser trochanter (most primates, not “australopiths”), a weakly
developed linea aspera (Orrorin), a femoral “carrying angle” (“australopiths” and
“Homo”), a tibia that is much shorter than the femur (at least apes), poorly
differentiated tibial tubercles (most primates), a medial tibial facet for the femoral
condyle lower than the lateral (at least apes), and a convex medial tibial facet (at
least apes) (cf. text and illustrations in Brown et al. 2004).

Is this specimen—if the cranial and postcranial remains represent a single
individual—a hominid? A gut reaction based on the external morphology of the
skull is “yes,” but the internal morphologies are odd. The teeth are not necessarily
hominid, the humeral shaft lacks the torque or “twist” characteristic of large-
bodied hominoids (Morwood et al. 2005) and the morphology of the distal
articular region is clearly not hominoid (cf. Schwartz 1986; Morwood et al.
2005), the tibia is definitely not hominid, the femur is Orrorin-like, and the
partial os coxa is somewhat “australopith”-like. So why is the composite Flores
specimen considered a species of Homo when its affinities to a clade Hominidae
are not entirely clear? Largely because there has been a history of allocating
specimens to taxa based more on their geological age than on their morpholo-
gy—which, in turn, has led to the general practice of trying to explain away
“anomalous” morphologies in terms of variation. Methodologically, however,
before one even contemplates referring a specimen to a genus and species, one
should have to defend first why any specimen is a hominid and then a member of
the smaller subclades that subsume that species. But in order to do so, we must
have a working definition of this potential clade that is open to criticism and
revision. In this regard, there is still much work to be done.

Acknowledgments

Drs. Henke, Rothe, and Tattersall certainly provided a challenge by asking me to
tackle this problem. In the end, it was an enlightening endeavor. But it would have
been impossible were it not for my long-term study with Ian Tattersall of much of the human fossil record, which depended on access to specimens (at the very least type specimens) afforded by most discoverers and museum curators in charge of specimens. Perhaps someday specimens from the Middle Awash, Ethiopia, and Chad will be generally available for study.

References


Blumenbach JF (1969) On the natural varieties of mankind (De Generis Humani Varietate Nativa) (1775 and 1795 volumes). Bergman, NY


Darwin C (1871) The Descent of Man, Parts I and II. John Murray, London


Defining Hominidae


Gregory WK (1922) The origin and evolution of the human dentition. Williams and Wilkins, Baltimore


Huxley TH (1863) Man’s place in nature. D Appleton, NY


Linnaeus C (1735) Systema naturae. Leiden


Maresca B, Schwartz JH (2006) Sudden origin: a general mechanism of evolution based on stress protein concentration and rapid environmental change. The Anatomical Record (New anatomist) 289B: 38–46


Schwartz JH (1999a) Evolutionary provocations: Paul Sondaar, the evolution of the horse, and a new look at the origin of species. Deinsea 7: 287–300


4 Role of Environmental Stimuli in Hominid Origins

Elisabeth S. Vrba

Abstract

Environmental stimuli have influenced the evolution of hominids and other mammals at the levels of ontogeny, organismal adaptation, and speciation. I review some agreement which has emerged—as well as persistent debates—on the issue of environmental linkages to hominid adaptation. I discuss some current hypotheses which link physical change, adaptation, and speciation in general and in hominids in particular (including hypotheses on the role of ecological specialization and generalization, the coordinated stasis and variability selection hypotheses, habitat theory and the turnover pulse hypothesis); and revisit some persistent debates (such as on whether or not there was mammalian species’ turnover in the Turkana Basin during the Plio-Pleistocene). The relation of hominid evolution to the recent finding of several turnover pulses coincident with global cooling trends in the 10-Ma Recent record of all African larger mammals is considered. One example of hypotheses which address issues of environmental stimuli of ontogenetic evolution is the heterochrony pulse hypothesis: the generative properties shared among lineages can result not only in coherence of morphological changes but also in a strongly nonrandom timing of heterochrony events, as diverse lineages respond in parallel by similar kinds of heterochrony to the same environmental changes. I give some mammalian and hominid examples involving body enlargement by prolongation of growth, and attendant “shuffling” of body proportions including relative increase in brain volume, namely encephalization.

4.1 Introduction

It is a truism that environmental stimuli have influenced the evolution of hominids and all other life forms. The challenge is to understand the causal subcategories: what are the hypotheses and predictions that should be tested, and what kinds of data can be used to best effect? My approach is based on three premises: (1) we need to study not only the hominids but also their wider biotic
and environmental context; (2) given the aim to understand hominid evolution, the theory of evolution should be accorded more prominence than has been the norm; and an expanded theoretical framework is needed. A focus on the dynamics which link the environment to selection and adaptation at the organismal level is insufficient. One also needs to consider the causal linkages from the environment to dynamics at lower and higher levels—from morphogenesis during organismal ontogeny to the macroevolutionary level of species turnover (speciation and extinction)—and investigate the separate and combined roles in the origins of new phenotypes and species; and (3) the direct influence of physical environmental stimuli on evolution at each level deserves more intensive study than it has been accorded traditionally. For much of the century following Darwin (1859), the research disciplines of geology (including climatology) and evolutionary paleontology were conducted separately. Speculations abounded on how they might link, but analyses directly integrating data from both areas remained sparse. This changed over the past 40–50 years as more refined methods led to discovery of new patterns and causal principles in paleoclimatology (e.g., the astronomical climatic cycles [Hays et al. 1976]) and in the fossil record (e.g., rigorous phylogenetic hypotheses, geochemical inference of past diet, etc.). Early papers linking evidence of physical change with hominid evolution concerned particular stratigraphic sequences in South and East Africa (respectively Vrba 1974, 1975; Coppens 1975; both compared climatic indications from mammalian change with the hominid record), the circum-Mediterranean area (Hsü et al. [1977] implicated the Messinian Salinity Crisis in hominid origin), and comparison of global climatic data with the hominid record (Brain 1981).

Darwin (1859) argued that the initiating causes of phenotypic change and speciation are located at the level of organisms, namely natural selection, particularly arising from competition: “... each new species is produced ... by having some advantage over those with which it comes into competition ...” (p 320). He stressed climatic effects on competition rather than on population structure: “in so far as climate chiefly acts in reducing food, it brings on the most severe struggle between the individuals ...” (p 68). Darwin thought that an understanding of organismal selection and adaptation will also answer the question of species’ origins. Most later evolutionary studies, including those in paleoanthropology, have continued in this tradition. (See Tattersall [1997a] who argued that “paleoanthropology fell completely under the sway of the evolutionary views of the [neoDarwinian] Synthesis—where it remains, for the most part, today.”) The questions I will address, using theory and evidence, include how physical dynamics have influenced species’ structure and speciation, organismal ontogenetic systems, and selection and adaptation. I also consider some of the interactions among these levels.
“Hominidae” is here understood in the traditional sense (Howell 1978) to exclude *Pan* and *Gorilla*. I use the term “species” for a sexually reproducing lineage, the members of which share a common fertilization system; and “speciation” for the divergence of the fertilization system in a daughter population to reproductive isolation from the parent species (Paterson 1982), with awareness of the difficulties involved in applying such concepts to the fossil record (Kimbel 1991; Vrba 1995a). The terms “habitat” and “habitat-specificity” of an organism or species refer to the set of resources that are necessary for life; resources are any components of the environment that can be used by an organism in its metabolism and activities, including temperature, relative humidity and water availability, substrate characteristics, places for living and sheltering, all kinds of organic foods, such as plants and prey, mates, and other mutualist organisms in the same or different species (Vrba 1992). An organism’s biotic environment derives from other organisms and biotic interactions such as competition, parasitism, predation, and mutualism. “Physical change” refers to the global and local effects from extraterrestrial sources, including the astronomical climatic cycles, and from dynamics in the earth’s crust and deeper layers as manifested by topographic changes such as rifting, uplift, sea level change, and volcanism.

### 4.2 Physical change, adaptation, speciation: some current hypotheses

The traditional hypothesis follows Darwin closely and has often been called neo-Darwinian. In its most conservative form, it assumes that adaptation and speciation are always driven by natural selection. The particular causes of selection are seen as very diverse, the most important being organismal interactions—such as competition and predation—that can act alone, or in combination with physical change, to initiate and complete speciation (and extinction). Under this null hypothesis, $H_0$, selection pressures that cause speciation differ from group to group and from one local area to the next. To explore how this model’s predictions differ from others, let us ask: what rhythm of speciation events would we expect if we could see all the events in the real world across the entire area under study (e.g., Africa) and if we plotted their frequencies against time? $H_0$ predicts that the pattern of origination frequencies for large areas, over long time, is a random walk in time with an averagely constant probability of origination. Examples of such arguments are found in Van Valen (1973), Hoffman (1989), McKee (1993), and Foley (1994). In contrast, a number of hypotheses share the argument that physical change is required for initiation of turnover, with the consequent prediction that speciation and extinction frequencies should be
nonrandomly distributed in time in association with episodes of physical change. I will discuss several such hypotheses after introducing some relevant theory.

4.2.1 Allopatric speciation

Allopatric speciation occurs in isolated populations that have been separated by vicariance or by dispersal over barriers. (Vicariance is the fragmentation of a formerly continuous species’ distribution into separated populations.) Gulick (1872), who studied the Hawaiian fauna, was the first to argue that the causes of speciation are not well explained by selection among competitors but that vicariance brought about by physical changes was seminal in initiating speciation. Mayr’s (1942, 1963) comprehensive arguments for allopatric speciation eventually resulted in widespread agreement that this mode predominates. Although there continue to appear claims of sympatric speciation, mainly in herbivorous insects and fishes, most recent such reports acknowledge that the best evidence remains circumstantial (review in Vrba 2005). It is fair to say that an expectation of predominant allopatric speciation, particularly in hominids and other large mammals, is consistent with the weight of available evidence and enjoys widespread consensus. In terms of earlier concepts of hominid phylogeny, which accepted a progression from the earliest biped to *Homo sapiens* with minimal branching from that lineage, one might wonder whether it is worthwhile to test causal hypotheses of hominid speciation; but recent finds indicate that “any accurate view of ourselves requires recognizing *Homo sapiens* as merely one more twig on a great branching bush of evolutionary experimentation” (Tattersall 1999 p 25, 2000). That is, we need to consider seriously the question of what caused lineage branching in the hominid tree.

4.2.2 Physical change as the driver of vicariance, selection, and speciation

If allopatric speciation predominates then so also must physical initiation of speciation. Vicariance is nearly always produced by tectonic and climatic change. Incipient speciation initiated by dispersal over barriers also in most cases implies the causal influence of physical change (e.g., chance *Drosophila* fly dispersals over the ocean always occur, whether there are islands within reach or not; it took the production of the precursor islands of the Hawaiian Archipelago for the founding of those first allopatric populations of Hawaiian drosophilids [Carson et al. 1970]). Thus, the chief causes of population size reduction and allopatry in the
history of life have probably derived from physical changes. Although the relationship of punctuated equilibria to physical change was not explored in Eldredge and Gould (1972), the pattern they argued for implies independently that the initiation of speciation mostly comes about through physical change (Vrba 1980): if species are in equilibrium for most of their durations, what causal agency of the punctuation can one invoke other than physical change? The general consensus on the importance of allopatric speciation, together with the implications of punctuated equilibria and Paterson’s (1978) “recognition concept” of species, led to the proposal that physical change is required for most speciation (Vrba 1980). Paterson (1978, 1982) argued that change in the fertilization system, the critical evolutionary change in sexual speciation, is most likely to occur in small, isolated populations that are under selection pressure from new environmental conditions.

4.2.3 Ecological specialists and generalists

This contrast is of general evolutionary interest (Stebbins 1950; Simpson 1953; Eldredge 1979; Vrba 1980, 1992) and is particularly germane to mammalian evolution during the climatic instability of the Late Neogene (Vrba 1987a). It has been discussed using various terminologies, such as stenotopy and eurytopy (Eldredge 1979), niche breadth (Futuyma 1979), and breadth of habitat specificity or resource use (Vrba 1987a). Specialist and generalist adaptation can be expressed in relation to different kinds of environmental variables, e.g., with respect to food intake, temperature, vegetation cover, light intensity, etc. Given the effects of the Late Neogene climatic and tectonic changes in Africa (see Sections 4.3.1 and 4.3.2), the distinction between species which are stenobiomic (restricted to a particular biome) and eurybiomic (ranging across biomes) is particularly relevant. As populations of a species encounter a new environment, beyond the ancestral biome range, they could in principle either diverge from their adaptation to the ancestral biome to become specialized on the new one, or become more eurybiomic by broadening their resource use to include the new biome alongside the ancestral one (Vrba 1987a, 1989a). Evolution toward eurybiomy, which is very rare (Hernández Fernández and Vrba [2005] found a large preponderance of biome specialists among living African mammals), is of special interest as it applies to Homo (Vrba 1985a, 1989a; Pickford 1991; Potts 1998; Wood and Strait 2004). Proposals that temporal and/or spatial environmental variability, namely life in fluctuating or unpredictable environments, can promote generalist adaptations have a long history of extensive discussion (Stevens [1989] reviewed the evolution of broad climatic
tolerance in high-latitude environments that have a greater range of annual and longer term variation). Adaptations to strong seasonality range across life forms, from diatoms and other photosynthetic groups in polar waters (which each winter form resting spores in response to darkening, sink down dormant out of the plankton environment to germinate again when light returns [Kitchell et al. 1986]) to the deciduous habit of many plants and hibernation and long-distance migration in animals. In advanced vertebrates, complex behaviors form an important category of such adaptations, ranging from the behavioral adjustments of animals to changing temperature and aridity (Maloyi 1972) to hominine culture.

The notion of a “biome generalist species” can be subdivided as follows. The eurybiomic phenotypes can be either (A) heritable, namely genetically based and fixed, or (B) expressed as ecophenotypes, within a broad norm of reaction, in response to varying environments (Hall 2001; West-Eberhard 2003). Case (A) has two subcases. (A1) Each organism can live in more than one biome either because each organism has the needed biomic flexibility or because each organism is a specialist on resource patches which occur across biomes. An example of the latter is the aardvark, *Orycteropus afer*, which is stenophagous (it eats only ants and termites), a substrate specialist (it digs burrows in sandy or clay soil), and stenophotic (it is nocturnal). Yet *O. afer* is eurybiomic: its specialized “resource patches” range from semidesert to dense, moist woodland across Africa. (A2) There are intraspecific differences in resource use among organisms and populations, i.e., polymorphism in resource use allows the species to respond to environmental fluctuations by shifting relative abundances of the variants. An example is the African buffalo (Sinclair 1977): *Syncerus caffer caffer* differs in phenotype and resource use from (and lives at higher latitudes and/or altitudes with more grassland present than) the smaller, plesiomorphic phenotype *S. c. nanus* (in warm, forested regions). This species appears to have “rolled with the punches” of large and frequent climatic changes since the Late Pliocene mainly by changes in polymorphic frequencies. All generalist adaptations first evolve in populations, thus rendering the species polymorphic. Intraspecific adaptations and polymorphisms, which become more elaborate with repeated climatic shifting, are likely to be the most frequent responses to climatic extremes (Vrba 1992: Fig. 4c) with speciation a much rarer outcome. Living African mammals include a few species which can live across several biomes from arid, open to well-watered, mesic, and closed environments (Hernández Fernández and Vrba 2005), i.e., their habitat specificities allow very broad tolerance of environmental variation over geography and through time. Thus, the notion that each species is “specific for a particular habitat” does not equate with specificity for a particular, or single, type of environment. The term “specific” here refers to the close relationship between species and their habitats, with no connotation of environmental
breadth. A species’ habitat may remain intact although it lives in strongly fluctuating environments over long time, such as the aardvark whose habitat and resources persisted as widely varying environments swept over the areas in which it lived.

### 4.2.4 Habitat theory and the turnover pulse hypothesis

#### 4.2.4.1 The Turnover Pulse Hypothesis

The Turnover Pulse Hypothesis is a part of the broader “habitat theory” which focuses on species’ habitats and on the dynamic relationships between physical change, habitats, and species (Vrba 1992). It uses the predominance of allopatric speciation and the consequence that physical change is required for most speciation. Climatic changes (from global or/and local tectonic sources) result in removal of resources from parts of the species’ former geographic distributions and therefore in vicariance. Vicariance on its own is insufficient for speciation. Many species underwent repeated episodes of geographic shifting, vicariance, and reunion, of their distributions (distribution drift), in response to the astronomical climatic oscillations, without speciation although intraspecific adaptive changes may have accumulated (Vrba 1992, 2005). For speciation to occur, physical change must be strong enough to produce population isolation but not so severe as to result in extinction; and the isolated phase must be of sufficiently long duration for the changes which define speciation (adaptation of the fertilization system to the new environment; Paterson [1982]) to occur. I have suggested that most speciation requires sustained isolation, or near isolation, without rapid reintegration on the Milankovitch timescale and that shrinking populations are important in which habitat resources are dwindling, competition increases, with consequent strong selection from the changing environment (Vrba 1995b). In the absence of physical change of appropriate kind and duration, although species may accumulate new adaptations, they are buffered against speciation at several levels (review in Vrba 2005).

One prediction is that most lineage turnover, speciation and extinction, has occurred in pulses, varying from tiny to massive in scale, across disparate groups of organisms, and in predictable temporal association with changes in the physical environment (Vrba 1985b, 2005). If we think of origination, several possible patterns of origination frequency could result, all different from the temporally random pattern predicted under $H_0$: (1) Origination could in principle be confined to rare, large pulses in response to the largest environmental changes. Such large pulses may resemble jagged mountain crests, or dissected high plateaus, rather than simple, single peaks because the timing of turnover
responses to climatic or tectonic episodes will differ among organismal groups and local areas. (2) Many, frequent, small pulses, such as in response to the 100 thousand year astronomical cycle, interspersed by the less frequent, larger ones described under (1). (3) Combinations of the random null model and the turnover pulse hypothesis suggest additional predictions such as a random background of turnover frequency punctuated by rare pulses. A comparison among turnover pulses is expected to show much heterogeneity—or “mosaic” differences. The environmental changes that trigger turnover are diverse. They vary in nature, intensity, timing—how long they endure, how much fluctuation occurs, and steepness of component changes and net trends—and in geographic emphasis and extent, from very localized to present in many parts of the earth. Topographic and latitudinal factors contribute to geographic variation in the turnover responses to a major global change. Also, the different organismal groups differ sharply in how they are affected by climatic variables (see Andrews and O’ Brien 2000 for mammals). They differ in turnover response (by speciation, extinction, or by no turnover at all). Lineages which do undergo turnover initiated by a given physical change may do so with different timing (in “relays,” see examples in Vrba 1995b, c). Thus, if a turnover pulse is detected in a data set, it is desirable to study subdivisions of those data to understand the detailed taxonomic, geographic, and temporal patterns.

4.2.4.2 Additional hypotheses

(1) Under habitat theory and other concepts which invoke predominant allopatric speciation, species should generally “start small,” namely in geographic distributions that are more restricted than those they attain later on (Vrba and DeGusta 2004). \( H_0 \) does not predict this. (2) Of two areas of similar large size, both subject over the same time to climatic cyclic extremes that remain habitable for organisms, the area that is more diverse in topography will have higher incidences of selection pressures and vicariance per species. The prediction is that the topographically more diverse area has higher rates of vicariance, speciation, and extinction (Vrba 1992). (3) During periods of strong latitudinal thermal contrasts, with ice caps on one or both poles, biomes closest to the equator are predicted to have higher speciation and extinction rates than biomes at adjacent, higher latitudes (e.g., this bias may have contributed to the high species richness in the tropics today [Vrba 1985b, 1992]). (4) Biome generalists are expected to have lower rates of vicariance, speciation, and extinction than biome specialists (Vrba 1980, 1987a, 1992). Because habitat theory stresses physically initiated vicariance and selection pressure, changes in amplitude and mean of the climatic
cycles, and in which cycle predominates, are expected to affect the evolutionary outcome. The larger the amplitude, the higher the incidence of vicariance and selection pressure at any cyclic extreme, accelerating the rates of intraspecific adaptation, speciation, and extinction. Changes in cyclic dominance can affect the frequency and duration of vicariance. Large translations in the climatic mean and envelope may be especially significant for speciation and extinction (Vrba 1995b: Figs. 3.2, 3.3).

4.2.5 The coordinated stasis hypothesis

Brett and Baird’s (1995) hypothesis of coordinated stasis is Darwinian in its focus on organismal interactions in a community as a source of stasis. It proposes that the coevolutionary bonds during stasis are so strong that physical change is needed to disrupt them to result in turnover. Thus, this model is “community based” in its theoretical assumptions (see also the hypothesis of coevolutionary disequilibrium of Graham and Lundelius 1984; and reviews in Barnosky 2001; Vrba 2005). Brett and Baird’s (1995) model predicts stasis of species, interrupted by pulses of speciation and extinction, across all communities in which a set of species occurs. Thus, their predictions are closely comparable with those of the turnover pulse hypothesis, as acknowledged by Brett and Baird (1995 p 287): “The same term [coordinated stasis] could be used for the blocks of stability in Vrba’s (1985) ‘stability-pulse’ hypothesis.”

4.2.6 The variability selection hypothesis

This hypothesis was proposed and applied to hominid evolution by Potts (1998; pages quoted are from this paper). It is about (1) a particular category of adaptations which he calls “variability-selection adaptations” (VSAs), (2) their initial appearance and establishment, (3) their fate in the face of long-term climatic cycles, and (4) an interpretation of the theoretical implications of their evolution.

1. VSAs are “structures and behaviors responsive to complex environmental change” (p 81), which are uniform within species “yet able to mediate secondary phenotypic traits that vary…” (p 85). His examples include a locomotor system allowing a wide repertoire of movement and “a large brain or specific neurological structure that is effective in processing external data and generating complex cognitive responses” (p 85).

2. VSAs arise first in isolated populations. Intraspecific polymorphism results with VSAs in some populations and not in others. Organismal selection
from short-term variability during organismal lifetimes initially promotes such VSAs (or at least allows them to persist).

3. The long-term evolutionary outcome at Milankovitch and longer timescales is that organisms with VSAs survive climatic extremes. Therefore, species which include at least one VSA-carrying population survive. Over time the VSAs can become more elaborate as climatic extremes recur. Thus, high climatic amplitude at timescales longer than organismal life times, notably at Milankovitch and longer timescales, causally influences the evolutionary outcome.

4. Climatic variability at the longer timescales is a selective agent of VSAs, which are “designed [by selection] to respond to novel and unpredictable adaptive settings” (p 85). That is, these organismal adaptations are shaped by selection for the function of flexible responses to future climatic excursions of the Milankovitch cycles; and this is a new kind of selection: “variability selection” (VS).

Potts claims that this hypothesis differs from all others, notably from the (p 82) “savanna hypothesis” and other “environmental hypotheses of hominid evolution [which focused] on a specific type of habitat.” He regards the VSA concept as distinct from previous concepts of adaptation, such as the generalist adaptations which confer eurytopy (Eldredge 1979), broad habitat specificity (Vrba 1987a), and broad climatic tolerance (Stevens 1989). He considers “habitat-specific” adaptations (Vrba 1987a) and selection pressures as different from VSAs and VS because the former in his view narrowly refer to a particular kind of environment and not to variable environments. I believe that he is wrong in these claims. Proposals (1), (2), and (3) are severally and jointly consistent with previous theoretical proposals (see Section 4.1.3). The sole departure is proposal (4). Structures and behaviors that confer flexibility in the face of climatic variations, and that may arise and exist as polymorphic variants in species, are well known (e.g., the resting/vegetative life cycle in diatoms noted above, hibernation, etc.) including complex behaviors in primates (e.g., the presence in some Japanese macaque populations of grass-washing behavior [Nakamishi et al. 1998]). But these adaptations require no more than selective agents during the life times of organisms. While they might fortuitously “come in handy” when the next ice age arrives, we would not claim that they must have been selected for that function. (See Williams’s [1966] argument that we should not confuse selection [which “cares” only about immediate fitness of the selected] and adaptation [which is no more than the character shaped directly by selection for current function] with the incidental evolutionary effects of these phenomena; see also Gould and Vrba’s [1982] distinction between adaptation and exaptation.) Also,
Potts (1998) is wrong in concluding that the “habitat-specific” adaptations and selection pressures, as under habitat theory, refer narrowly to a particular kind of environment and not to variable environments. As noted earlier, “habitat” should not be conflated with “environment.” A species’ habitat may remain intact in different environments at one time and over a long time such as the aardvark whose habitat and resources persisted as strongly differing environments swept over the areas in which it lived.

The special effects which the high amplitude of the climatic cycles since the Late Pliocene had on the biota (e.g., that species, in which generalist adaptations for climatic tolerance had evolved survived disproportionately) have also been discussed (Stanley 1985; Vrba 1985a, 1992, 2000). No one doubts that strong Milankovitch excursions can selectively remove some populations and species whose organisms are unfit under those conditions or that generalist adaptations of survivors can sequentially be elaborated during recurrent such episodes. But this would not be a new kind of selection. Organismal selection cannot promote adaptations to future Milankovitch extremes, although “inter-demic selection” (Wright 1932, 1967) or species selection (review in Vrba 1989b) could in principle occur at those longer timescales. (In fact, Potts did at one stage wonder whether his notion might represent a form of lineage or species selection [R. Potts personal communications]). The problem is that the concept of selection and adaptation at levels higher than that of organisms is onerous (Williams 1966; Vrba and Gould 1986; Maynard Smith 1987; Vrba 1989b). Maynard Smith (1987) discussed this as follows (p 121): “We are asking whether there are entities other than [organisms] with the properties of multiplication, heredity, and variation, and that therefore evolve adaptations by natural selection.” Considering the nature of the adaptations Potts (1998) had in mind, one probably does not need to invoke higher level selection. Such issues on levels of selection have been extensively debated and with respect to diverse organismal case histories. An example which is of interest here, in spite of (and perhaps because of) being far removed from hominids on the tree of life, is the case in diatoms with the resting stage adaptation to polar conditions of long winter darkness (Kitchell et al. 1986): The fossil record shows that diatoms living in Arctic waters just before the Cretaceous/Tertiary (K/T) boundary already had this life cycle adaptation. Kitchell et al. (1986) documented that during the K/T mass extinction (which involved long-term global darkening), diatoms and other photosynthetic planktonic groups with resting stages had markedly lower extinction rates than groups which lacked this seasonal adaptation. They argued (correctly in my view) that these life history features, which arose by selection at the organismal level as adaptations to seasonal variability, were also fortuitously (by sheer luck) available and useful during the K/T event for weathering much longer intervals of darkness.
They concluded that this sorting among species, although nonrandom, does not represent species selection but species sorting according to the effect hypothesis (Vrba 1980). The adaptation in this case could not have been selected at the organismal level for climatic variability at the timescale of mass extinction. I suggest that the selective forces and character complexes which contributed to survival of hominids and other mammals in the face of increasing climatic amplitude during the Late Neogene may in principle fall into the same category. That is, no new kind of selection needs to be invoked.

4.2.7 Tests based on the temporal distribution of newly appearing phenotypes and species

Both the coordinated stasis and turnover pulse hypotheses predict significant concentration in time of speciation and extinction events (namely, turnover pulses). The prediction of Potts’s (1998) model is that features that enhance flexibility should appear “during a period of expanding environmental fluctuation” (p 92). Most difficulties in testing such hypotheses have to do with errors in the chronological, physical, and biotic data, and with testing at inappropriate temporal, geographic, and taxonomic scales (Barnosky 2001). Vrba (2005) discussed two types of error in tests for turnover pulses: inferring pulses that are not there (i.e., erroneous rejection of $H_0$, type I error), and failure to detect real pulses (type II error) as exemplified by Signor and Lipps (1982). Take for an example of type II error an attempt to distinguish between $H_0$ and real small speciation pulses that occurred at the Milankovitch timescale. Such a test using first appearance data with lower time resolution (e.g., the data for intervals of 0.5 Myr [million years; Ma will refer to million years ago] length in Vrba and DeGusta’s [2004] study of African mammals) will fail. In most available data sets, the best hope lies in testing whether or not large turnover pulses can be detected. The main bias that leads to type I error, seeing pulses that are not there, arises from unequal fossil preservation between time intervals, areas, and groups of organisms, the “gap bias” (Vrba 2005): any given species’ fossil FAD (first appearance datum) may postdate its true, or cladistic, FAD (Kimbel 1995). Gaps have the effect that, for instance, a count of FADs in an interval is erroneously inflated by species’ records that in reality originated (but were not detected) previously. An early version of a test that corrects for the “gap bias,” thus allowing a rigorous test of the pulse hypothesis, was applied to the African larger mammals of the past 20 Myr divided into 1-Myr long intervals (Vrba 2000). More recently, a second updated form was applied to the nearly 500 species recorded over the past 10 Myr divided into 0.5-Myr long intervals (Vrba and DeGusta 2004; Vrba 2005). Time resolution in this record is sufficiently good, with
more than 70% of the site records dated by radiometric or paleomagnetic means, that any large speciation (or extinction) pulses spaced 1 Myr apart should be detectable. Some results will be mentioned in Section 4.3.3.

4.3 Physical change, adaptation, speciation: evidence from the African Neogene

4.3.1 Physical background: climatic change

Following the definitive documentation of the astronomical cycles (Hays et al. 1976), it was thought that they may have had little effect on the tropics in general (review in Burckle 1995) and on African hominid-associated environments in particular. For example, Hill (1995 p 187) considered that: “it may be that African terrestrial vertebrate habitats were to some extent buffered from climatic changes seen elsewhere.” Hill’s caution is well taken that specific areas may “march to a local drummer” especially if that drumbeat derives from tectogenesis (see below). It now appears that much of Africa participated in the global changes. The Late Miocene-Recent record is relevant as the earliest hominid fossils currently date to ca. 7 Ma (Brunet et al. 2004).

4.3.1.1 The Late Miocene

There was ice buildup on West Antarctica and general increase in δ¹⁸O values 7–5 Ma (Müller and Fairbanks 1985; Kennett 1995). A major cooling which started before 6 Ma and peaked shortly thereafter contributed to isolation and desiccation of the Mediterranean Basin during the Messinian low-sea-level event and salinity crisis dated ca. 5.8–5.3 Ma (Haq et al. 1980; Hodell et al. 1994; Bernor and Lipscomb 1995; Aifa et al. 2003; Garcia et al. 2004). This was followed by warming and a post-Messinian transgressive phase starting before 5 Ma to reach a maximum in the 5- to 4-Myr interval (Haq et al. 1987). Reviews of the numerous and often major changes—both physical and biotic—on the continents and in the oceans, and particularly in Africa, can be found in Brain (1981) and Vrba et al. (1995). Questions remain on the African effects of these Late Miocene climatic events. Kingston et al. (1994) found that, in the Kenyan Tugen Hills area, a heterogeneous environment with a mix of C₃ and C₄ plants—and without grassland dominance—persisted over the entire past 15 Myr without any apparent local influence from global climatic change. Evidence from Lothagam indicates that this part of Kenya experienced strong environmental changes over the latest
Miocene (Leakey et al. 1996; Leakey and Harris 2003). Further evidence comes from analyses of carbon isotope ratios in soils and fossil tooth enamel. Cerling et al. (1997) studied fossil herbivores ranging over the past 22 Myr from several continents. Using the fact that low $\delta^{13}C$ values in herbivore teeth reflect a diet of mainly C$_3$ plants while high values indicate feeding on C$_4$ plants, they found that up to 8 Ma, mammals in Pakistan, Africa, and South and North America had C$_3$ diets or C$_3$-dominated diets. By the late to latest Miocene C$_4$-plants came to dominate the diets. In Kenya, representing the lowest latitude in the sample, the transition was complete by between 8 and 6.5 Ma, and in Pakistan by ca. 5 Ma. Cerling et al. (1997) interpreted their results as showing a global increase in the biomass of C$_4$ plants between 8–6 Ma which resulted from a decrease in atmospheric CO$_2$.

4.3.1.2 The Plio-Pleistocene

Much evidence by now suggests that many parts of Africa during the Plio-Pleistocene were indeed responding to orbital variations and certainly to the strongest global events (Reviews in Vrba et al. 1995; Reed 1997; Potts 1998; deMenocal 2004). Continuous Plio-Pleistocene records which demonstrate this are now known from western, eastern, northern, and southern Africa: Pollen cores off West Africa record the shifting of the Sahara–Sahel boundary and the earliest extensive spread of the Sahara desert ca. 2.8–2.7 Ma (Dupont and Leroy 1995). Marine records off West Africa and from the Gulf of Aden document $\delta^{18}O$ variations, and also dust influxes from the Sahara and Sahel regions in the West, and from Arabian and northeastern African areas in the Gulf of Aden (DeMenocal and Bloemendal 1995; deMenocal 2004). A record of aeolian grain size from the Gulf of Aden shows changes in the intensity and phase of the Indian monsoon (Clemens et al. 1996). A northeastern African record shows that sapropels (organic-rich black layers) were deposited during humid periods in the eastern Mediterranean Sea following high-flood periods of the Nile River at the tempo of orbital precession (Rossignol-Strick et al. 1998). For southwestern Africa, a record of the past 5 Myr off Namibia, underlying the Benguela upwelling system, yielded a continuous time series of changing sea surface temperature (SST) for the past 4.5 Myr, with decreased upwelling interpreted to represent warmer conditions with wetter, more mesic periods in southern Africa (Marlow et al. 2000). A marine record off Angola derived from compound-specific carbon isotope analyses of wind-transported terrigenous plant waxes, indicated African C$_4$ plant abundances during the interval 1.2–0.45 Ma and showed that the African vegetation changes are linked to SST in the tropical Atlantic Ocean and that changes in atmospheric moisture content due to tropical SST changes and the
strength of the African monsoon controlled African aridity and vegetation changes (Schefuss et al. 2003). The phase relationships between the African monsoon and the glacial cycles were shifting continuously (Clemens et al. 1996), explaining why indicators of surface water, such as lake levels, and of vegetation, such as dust spikes, do not always covary (deMenocal et al. 1993).

The major events and steplike changes identified by these and other authors are as follows. There were significant shifts in the intensity and phase of the Indian monsoon at ca. 2.6, 1.7, 1.2, and 0.6 Ma (Clemens et al. 1996). Marlow et al. (2000) concluded that SSTs decreased markedly, in association with intensified Benguela upwelling, after 3.2 Ma, with subsequent periods of marked SST decrease and upwelling intensification near 2.0 and 0.6 Ma. DeMenocal and Bloemendal (1995) documented a shift from dominant climatic influence occurring at 23–19 thousand year periodicity prior to ca. 2.8 Ma to one at 41 thousand year variance thereafter, with further increases in 100 thousand year variance after 0.9 Ma (see also Ruddiman and Raymo 1988). DeMenocal (2004) summarized his conclusions on consistent patterns of African subtropical climatic variability, based on the West and East African records analyzed by him and others, as follows (p 8): “1. Orbital-scale climatic variability persisted throughout the entire [Plio-Pleistocene] interval extending in some cases into the Miocene. . . . 2. The onset of large-amplitude African aridity cycles was closely linked to the onset and amplification of high-latitude glacial cycles. 3. Eolian concentration and supply (flux) increased gradually after 2.8 Ma. 4. Step-like shifts in the amplitude and period of eolian variability occurred at 2.8 (±0.2) Ma, 1.7 (±0.1) Ma, and 1.0 (±0.2) Ma. 5. Evidence for 10^4–10^5 year ‘packets’ of high- and low-amplitude palaeoclimatic variability which were paced by orbital eccentricity.” Using records of δ¹⁸O variation since the Late Oligocene, Potts (1998) subtracted the lowest from highest value for each unit million year as a measure of total climatic variability. He found that variation of 0.3–0.5 parts per mil (ppm) was obtained for most of the Neogene until the 6.0- to 5.0-Myr interval, during which variability rose sharply. After a minor decrease during 5–4 Ma, there were increases during every succeeding interval with the highest one to 1.9 ppm, during the past 1 Myr (Potts 1998 p 83: Fig. 1).

I used Shackleton’s (1995: Fig. 17.3) data, which records δ¹⁸O variation at 0.003-Ma interval steps for the past 6 Myr, to identify periods over which the largest net cooling or warming trends occurred. (Vrba 2004: take an interval t_x of length x thousand year, for example t_{100} for x = 100 thousand year, and move it step by step along the time axis from early to late. At each interval step, mark the interval along the time axis if either of conditions C, for cooling, or W, for warming, is true: C: the upper [warm] envelope of the climatic curve remained continuously below the running mean of the previous 300 thousand year; i.e., the interval is a t_{100,C} or an interval of length 100 thousand year with marked
cooling. \(W\) is the corresponding condition for a warming trend. A pattern of \(t_{100,C}\) and \(t_{100,W}\) distribution in time results, with data clusters for the most sustained trends. I here report results for separate assessments using interval lengths in \(x\) thousand of year of \(x = 40, 65, 100, 140\).) The following approximate intervals (in chronological sequence) emerged as times of sustained net cooling (\(t_{40,C}, t_{65,C}, t_{100,C}, t_{140,C}\), are respectively labeled *, **, ***, ****, from least to most severe; time ranges in ca. Ma): 5.1–4.90*, 4.1–3.95*, 3.45–3.25***, 2.7–2.35**** (and 2.9–2.5*), 2.1–2.0*, 1.8–1.65**, 0.95–0.85**, 0.8–0.65** (and 0.8–0.6*). Intervals of net warming (similar notation as for cooling): 5.6–5.35****, 4.5–4.4*, 3.1–2.9*, 1.65–1.6*, and ca. 0.85*. Hypotheses under habitat theory predict that such strong longer-term climatic trends should be associated with higher frequencies of evolution in general and of speciation and extinction particularly. (The predictions of the variability selection hypothesis [Potts 1998] do not differ substantially from this: times of strong change in climatic mean are also those of increased variability as measured by Potts [1998]. For times of increased variability without strong change in climatic mean, both predict an increase in intraspecific adaptation including adaptation which confers increased flexibility in the face of the climatic extremes.)

The precise relationships between the different kinds of evidence cited are incompletely understood; and, as expected, time estimates for significant climatic events do not neatly coincide. We are far short of having a set of “paleoclimatic magic numbers” (Lewin 1984 p 154) for comparison with hominid evolution. Nevertheless, it is worth noting the rough consensus on periods of major change: during the latest Miocene these included that toward 6.5 Ma and that associated with the Messinian ca. 5.8–5.3 Ma; but the African effects of changes over this general time period are less well defined because of a paucity of long records. Several intervals during the Plio-Pleistocene stand out (in approximate descending order of magnitude in ca. million years): notably 2.9–2.3, 3.5–3.2, 1.8–1.6, 1.2 and 1.0–0.6, and possibly also ca. 5 Ma, toward 4 Ma, and ca. 2 Ma.

In sum, hominid and other lineages faced net increase in seasonal cooling, aridification, and vegetational opening (see Brain [1985] on how cooler winters promote reduced height and wider spacing of tree cover in Africa). These variables have strong effects on species distribution and richness today (e.g., Andrews and O’Brien [2000] found that plant species richness and thermal seasonality were the most important such factors in modern southern African mammals as a whole and in large mammals annual temperature as well). Resource availability was fluctuating at seasonal to Milankovitch timescales and at intermediate (Potts 1998) and longer timescales. While over the long term cooling and aridification were both increasing and many African records show that association, it is by no means invariable as expected from the shifting phase relationship between the monsoon and glacial
cycles. For example, pollen data from Hadar, Ethiopia, shows that “*Australopithecus afarensis* accommodated to substantial environmental variability between 3.4 and 2.9 Ma ago. A large biome shift, up to 5°C cooling, and a 200- to 300-mm/year rainfall increase occurred just before 3.3 Ma, which is consistent with a global marine δ^18O isotopic shift” (Bonnefille et al. 2004 p 12125).

### 4.3.2 Physical background: tectonism

Tectogenesis has featured less prominently than climate change in discussions of evolution, perhaps because it is mostly a slow process and date limits for events tend to be wide. Yet it has had a primary influence on landscape and biotic evolution. This includes hominid evolution especially in rift-associated environments as recognized long ago by Coppens (1988–1989). Crustal changes influenced climate on a grand scale, e.g., Late Pliocene closure of the isthmus of Panama may have led to the start of the modern ice age (Maier-Reimer et al. 1990; Haug et al. 2001). Uplift of western North America, the Himalayas, and the Tibetan Plateau, possibly influenced the Pleistocene cooling intensification at ca. 1 Ma (Ruddiman et al. 1986). Northward drift of Africa during the Neogene led to southward displacement and areal decrease of tropical African forests and contributed to long-term aridification (Brown 1995). Episodes of intensified African uplift since ca. 30 Ma, which raised the entire eastern surface higher than in the West, greatly affected the African climate (Burke 1996). Apart from the numerous localized climatic effects of tectogenesis (Feibel 1997), the topographic diversity it generates together with the superimposed climatic cycles constitutes a prime cause of spatial and temporal environmental heterogeneity, changing selection pressures and speciation (Vrba 1992). Thus, evolution of the African Rift had an especial role in some evolutionary events in hominids (Coppens 1988–1989) and other mammals (Denys et al. 1987). The present episode of rifting began in the Early Miocene (Frostick et al. 1986). Between Ca. 8 and 6 Ma, a general increase in African tectonic activity led to formation of the Western Rift (Ebinger 1989). A major episode of uplift coincided with the climatic changes ca. 2.5 Ma (Partridge and Maud 1987). After 6 Ma, the rift system continued to propagate to the southwest toward the Kalahari Craton (Summerfield 1996). One incipient zone of rifting, trending southwest from Lake Tanganyika, terminates in central Botswana, where faulting and tilting of the zonal margins have resulted in damming of the Okavango River to spread out as the extensive inland Okavango Delta (Scholz et al. 1976).

I suggest that the dynamics of the hydrological features associated with rifting – rivers redirected, lakes forming and disappearing, and especially the
inland deltas spreading at the margins of incipient rift zones – have had a particular impact on the evolution of hominids and other biota. All early hominids required permanent water, and many of the eastern African hominid sequences reflect riverine and rift-margin associated deltaic and lake environments (Harris et al. 1988; Brown and Feibel 1991). The significance of inland deltas is that they can form vicariated “islands” of mesic conditions—or refugia—throughout periods of aridification and even in the absence of topographic heterogeneity. The edges of such a refugium are ecologically heterogeneous with intrusions of the arid surrounding environment. (“Refugium” here means a biome refugium, e.g., a forest refugium preserves the characteristic forest vegetation physiognomy, although its detailed taxonomic composition may differ from that of the parent forest community.) The Okavango Delta provides a good example: it is a vicariant island—despite the very low relief of the area (Scholz et al. 1976)—of woodland savanna and water almost entirely surrounded by semidesert. Vrba (1988) suggested that many of the hominid-bearing strata represent times when the areas were such inland deltaic-riverine-lacustrine refugia and that this poses problems for our ability to recognize times of widespread climatic change across the larger areas because “climatic change in the larger region is recorded in a refugium only close to its ecotonal limits, by the new appearances (or disappearances) of peripheral taxa . . . that represent occasional intrusive elements from the alternative biome” (p 410). An important implication from the evolutionary perspective is this: as climatic changes were sweeping across much of Africa at the Milankovitch scale, so such inland deltas were recurrently isolated and reconnected as parts of larger continuous biomes. During the reconnected phases, migration and gene flow occurred. During the vicariant phases there was enhanced incidence of gene pool divergence among populations, selection pressures at the refugial margins, intraspecific phenotypic diversification, and speciation. If true that inland deltas can in this way act as centers of phenotypic diversity and speciation and that they are particularly prevalent at the tilting margins of incipient rift zones, this would predict a Late Neogene propagation of centers of increased speciation in a south-southwesterly direction as the rift evolved.

4.3.3 The record of first appearances of mammalian species

4.3.3.1 All larger mammals

As noted above, a method which corrects for the “gap bias” was applied to the African larger mammal record of the past 10 Myr. Such correction is especially important in the Late Neogene climatic context because open, mesic to arid areas
tend to preserve vertebrate fossils better than do the more forested, wetter ones (Hare 1980). The following results emerged (largely agreeing with those in Vrba [1995c, 2000], in so far as they are comparable): Over the past 8 Myr, the strongest turnover pulses, involving both origination and extinction, occurred in the 5.5- to 5.0- and 3.0- to 2.5-Ma intervals. (The division into equal time intervals is artificial. The dating of the earlier pulse is tentative and this event may be closer to 5.5 Ma, or just before, than to 5 Ma; I will refer to the 5.5-Ma event. The Late Pliocene changes are better delimited between ca. 2.8 and 2.3 Ma.) Each of the intervals 7.0–6.5 and 3.5–3.0 Ma had an origination pulse without an extinction pulse and 1.0–0.5 Ma an extinction pulse without an origination pulse. Where one can compare this set of turnover events with the strongest cooling trends, the coincidence in time and intensity is strikingly close: the strongest climatic event, cooling toward ~2.5 Ma, coincides with the strongest turnover pulse, while lesser cooling and turnover events are present in the intervals 3.5–3.0 and 1.0–0.5 Ma. The results also showed intervals of significantly low origination and extinction, some of which overlapped with periods of high sea level with low polar ice on a warmer earth (Haq et al. 1987; Hodell and Warnke 1991).

The African mammalian record, and the bias-correction model which was used, continue to be updated. The results do give preliminary support to the hypothesis that at least a substantial part of turnover in African mammals was initiated by climatic change and that global cooling with increased aridity and seasonality was a more important stimulus of turnover than was global warming (Vrba 2000, 2005). Of the cooling trends, the one toward 2.5 Ma was the strongest, followed by a lesser trend starting ~1 Ma. Yet individual glacial maxima became colder after 2.5 Ma, especially after 1 Ma (Shackleton 1995). The fact that there were no further major origination pulses after 2.5 Ma suggests that most of the lineages present then were either species that had evolved during the start of the modern ice age with adaptations to the new environments or long-lasting biome generalists that survived right through that cooling trend.

A related result is that of Vrba and DeGusta (2004). We studied the question whether most species “start small,” namely in geographic distributions that are more restricted than those they attain later on. We used the same 10-Myr long record of the African larger mammals and the correction for the “gap bias.” The number of fossil site records, from which each species is known in an interval, was taken as a proxy for the magnitude of its living geographic range and abundance in that interval. We then tested $H_0$ that the geographic spread of species remained averagely constant across successive survivorship categories, namely from the first appearance (FAD) interval to the immediately following one, and so on. We found that the mean number of site records increased strongly from the FAD interval to the following survivorship interval, followed by a less marked although
still significant increase to the next interval, with no significant changes thereafter. Thus, we concluded that the average large African mammal species has indeed started its life in a relatively small population, and thereafter increased in geographic range to reach its long-term equilibrium abundance by ca. 1 Ma after origin. This supports hypotheses of speciation that accord a major role to the formation of isolated populations of reduced size initiated by physical change.

Not everyone has agreed that global change was a driver of evolutionary change and speciation in African hominids and other mammals. Kingston et al.’s (1994) conclusion for the Tugen Hills area in Kenya that similar heterogeneous environments persisted over the past 15 Myr without any apparent local influence from major global climatic episodes is not necessarily at variance with my findings for all larger African mammals. It seems possible that the effects of tectonism in the Tugen Hills locally dominated any signal in the area from widespread climatic change. One aim of Behrensmeyer et al. (1997) was to test Vrba’s (1995c) finding of a turnover pulse in African mammals between ca. 2.8 and 2.5 Ma by examining the past 4.5 Myr in the Turkana Basin (including the northern Shungura Formation, Ethiopia, and the southwestern Nachukui and southeastern Koobi Fora Formations in Kenya). They concluded that there was “no major turnover event between 3.0 and 2.5 Myr” (p 1591) and that this “weakens the case for rapid climatic forcing of continent-scale ... faunal turnover” (p 1593). I have reservations about their methods and assumptions which differed substantially from mine (Vrba 2005). A reexamination of Turkana Basin evolution over 4–1 Myr divided into 0.5-Ma long intervals, using my African mammal data base and the statistical “gap bias” model outlined above, showed a single significant origination pulse in the 3.0- to 2.5-Ma interval and no extinction pulses (Vrba 2005). Separate examination of the northern and two southern areas of the Turkana Basin indicated a strong speciation (and extinction) pulse in the North 3.0–2.5 Ma, but none in the combined or separate southern areas. This result is consistent with the southward spread of the Sahara Desert in the latest Pliocene (Dupont and Leroy 1995), which affected the northern basin more strongly, eliciting significant turnover, while the southern deltaic-lacustrine areas may have behaved more nearly like biome refugia.

At least some studies show that the larger mammalian turnover pattern is also reflected in small mammals. Among micromammals of the Shungura Formation, Ethiopia (Wesslaman 1995), at 2.9 Ma, woodland taxa predominated and even rainforest taxa were present (e.g., the bushbaby *Galago demidovii*, a rainforest species today). These forms were displaced by new grassland-to-semidesert species by 2.4 Ma. The turnover includes terminal extinctions, immigrants from Eurasia, such as a hare, *Lepus*, and global first appearances of species such as a new species of *Heterocephalus*, the genus of desert-adapted naked molerats, and a
new species of the ground squirrel genus *Xerus* (Wesselman 1995). This time also marks the first African and global debuts in the record of several species of bipedal, steppe-, and desert-adapted rodents such as the genus *Jaculus* of desert gerboas (Wesselman 1995) and a new springhare species, *Pedetes*, in South Africa.

### 4.3.3.2 The hominid record

The hominid sample is too small (15 to more than 20 species depending on which sources are consulted) to test whether most hominid species “started small” and to test for turnover pulses using the statistical methods which were applied to all larger mammals. Nevertheless, it is of interest to compare the known hominid FAD record with the timing of major climatic trends and speciation pulses in all larger African mammals. The earliest appearance of hominids, *Sahelanthropus* from Chad (Brunet et al. 2004; ca. 7 Ma; the hominid clade originated 8–5 Ma based on molecular estimates, Ruvolo 1997), forms a part of the elevated mammalian origination during 7–6.5 Ma and falls in an interval marked by increased African tectonic activity (Ebinger 1989) and ice buildup in West Antarctica with global cooling (Kennett 1995). *Ardipithecus* is first recorded just before 5.5 Ma (Haile-Selassie 2001), close to the turnover event in mammals and the major cooling and regression events associated with the Messinian ca. 5.8–5.3 Ma. One may question whether climatic change was a factor in the origin of *A. ramidus* because the Aramis stratum in which it occurs seems to represent a wooded and perhaps forested environment (WoldeGabriel et al. 1994). But as argued below in Section 4.3.4, this may not preclude the initiation by physical change of speciation in this case.

Most African FADs of hominid species are Pliocene to mid-Pleistocene in age during which the intervals of strongest climatic change were (in roughly descending order of magnitude, in ca. Ma; see Section 4.3.1.2): 2.9–2.3, 3.5–3.2, 1.8–1.6, 1.2 and 1.0–0.6, and possibly also ca. 5 Myr, toward 4 Ma, and ca. 2 Ma. Together these cooling episodes occupy ca. 40% of the past 5 Ma. Yet most, and possibly all, of the hominid FADs either coincide with or fall very close to one of these cooling events (chronology after Wood and Richmond 2000): *Australopithecus anamensis* and *A. afarensis*, FADs ca. 4.2 Ma; FADs of *A. bahrelghazali*, *Kenyanthropus platyops*, and possibly also *A. africanus* are a part of the origination pulse in the 3.5– to 3.0-Ma interval; *Australopithecus garhi*, *Paranthropus aethiopicus*, *P. boisei*, and possibly also *Homo habilis* and *H. rudolfensis* have FADs in the 2.8- to 2.3-Ma interval. FADs of *H. ergaster*, *H. rudolfensis*, and of *H. erectus* (and its migration to Eurasia) fall between 2.0 and 1.8 Ma. While taphonomic factors and chance may have contributed to this pattern, it does leave intact the hypothesis of climatic
cause of hominid speciation. An important splitting event in the hominid clade was the one that led to *Paranthropus* on the one hand and *Homo* on the other. Several systematic studies have concluded that the characters of *A. afarensis* are consistent with it being the common ancestor of *Paranthropus* and *Homo* and possibly also of one or more additional lineages (Kimbel 1995; Asfaw et al. 1999). After enduring in apparent equilibrium since at least 4.2 Ma, *A. afarensis* is last recorded just after 3.0 Ma (Kimbel et al. 1994), while its descendants appear variously between 2.7 and 2.3 Ma. Kimbel (1995 p 435) concluded: “regardless of which phylogenetic hypothesis is more accurate, it is clear that a pulse of speciation occurred in the hominid lineage between 3.0 and ca. 2.7 Ma, producing at least three lineages.” The phylogenetic pattern, of an inferred ancestor ending near 2.9, with new descendants branching off between 2.9 and 2.3 Ma, is common in bovids (Vrba 1995c, 1998a). These concordant genealogical patterns among different mammalian groups strongly suggest the causal influence of the start of the modern ice age, namely that common causal rules connect the climate system with evolution of different biotic groups. It remains to be seen whether additional information in future will support these preliminary indications that major climatic cooling trends and the concomitant changes in African environments were important causal influences on speciation in Hominidae, just as they were in many other mammalian lineages.

### 4.3.4 Climate and adaptation

There is widespread agreement that many of the Late Neogene mammalian morphological and behavioral changes, including in hominids, are overall broadly consistent with the extraordinary climatic fluctuations, and net cooling and aridification, over that time. But the precise nature of the environmental stimuli and of the processes they set in motion remain elusive. Nowhere is this more true than in the case of the origin of human bipedalism. According to Harcourt-Smith and Aiello (2004), the earliest evidence for bipedalism is arguably from *Sahelanthropus tchadensis* dating to 7–6 Ma (Brunet et al. 2002), *Orrorin tugenensis* ca. 6 Ma (Senut et al. 2001), and *Ardipithecus ramidus kadabba* first recorded just before 5.5 Ma (Haile-Selassie 2001). Wood and Richmond (2000) considered the tibia of *Australopithecus anamensis* (ca. 4.2 Ma; Leakey et al. 1995), the earliest undisputed evidence of bipedalism. Morphological evidence of a commitment to long-range bipedalism (e.g., long legs, large femoral head) appeared much later, ca. 1.6 Ma, in the postcranial skeleton KNM-WT 15000 from Nariokotome, West Turkana (Brown et al. [1985], who assigned it to *H. erectus*; Wood and Richmond [2000], included it in *H. ergaster*). The onset of this morphology in the fossil
record falls into one of the periods of strong climatic change toward seasonal aridification and a greater prevalence of open vegetation. The ecological and evolutionary implications of the bipedal record before this time remain a matter of unresolved debate. Harcourt and Aiello (2004) reviewed the evidence (including the Laetoli footprints, the AL 288-1 *A. afarensis* skeleton, postcranial material from Koobi Fora, the Nariokotome *H. ergaster* skeleton, “Little Foot” [Stw 573] from Sterkfontein, South Africa, fossils of *Orrorin, Ardipithecus*, and *Sahelanthropus*) and proposed a greater diversity in bipedalism in earlier hominids than previously suspected. Such independent evolution of different styles of bipedalism within the hominid clade by itself suggests that environmental change was a cause. In each of their three phylogenetic scenarios (Harcourt-Smith and Aiello 2004: Fig. 4), the postcranial diversification coincides broadly with the 3.0- to 2.3-Ma period of the largest Pliocene climatic trend. Stanley (1992) accepted that species of *Homo* were the first hominids to show adaptations to open arid environments and argued that climatic change was the cause. He hypothesized that the earlier gracile australopiths did not have expanded brain sizes due to their semiarboreal mode of life which necessitated that neonates be mature enough to cling to climbing mothers. The onset of an increased encephalization quotient (EQ: based on brain volume corrected for body weight) required that neonates be born in a more altricial state and unable to cling to mothers. Thus, only once semiarboreality was abandoned could encephalization evolve.

It had long been accepted that Late Miocene African environmental change, from widespread forest to more open vegetation, resulted in new selection pressures and thus set the stage for the origin of bipedalism (the “savanna hypothesis”). Specific hypotheses of what caused the adoption of upright posture in such a context have included (reviews in McHenry 1982; Preuschoft 2004): carrying, display or warning, new feeding adaptations, control of body temperature, tools, stone throwing, and wading in shallow water. McHenry (1982) thought that hominid bipedalism “could have arisen as an energetically efficient mode of terrestrial locomotion for a small-bodied hominoid moving between arboreal feeding sites” (p 163). Recent paleoecological findings for the earliest hominids have led to doubts of the savanna hypothesis. For example, Pickford et al. (2004) found a water chevrotain *Hyemoschus aquaticus*, indicating dense rainforest, in the Mabaget Formation, Kenya, dated 5.3–4.5 Ma, and associated with a hominid species. The evidence associated with *Ardipithecus* at Aramis in Ethiopia indicates a relatively closed, tree-dominated habitat (WoldeGabriel et al. 1994). Clarke and Tobias (1995) proposed that the foot bones from Sterkfontein Member 2 (Stw 573, Little Foot, dated ca. 3.5–3 Ma) reflect a foot that had not sacrificed arboreal competence or hallucial opposability and that this suggests dense tree cover in the environment. Based on fossil pollen, it has been suggested
that the preferred habitat of *A. africanus* at Makapansgat was subtropical forest and that selective pressures associated with densely vegetated environments played a role in the evolution of bipedalism (Cadman and Rayner 1989; Rayner et al. 1993). Potts (1998) dubbed this the “forest hypothesis” of bipedal origin and countered it by pointing out that most of the Late Miocene–Pliocene hominid species appear to have lived in varied habitats ranging from closed to more open conditions. The fossil bovids associated with *A. africanus* at Makapansgat and Sterkfontain do not suggest a uniform forest, although a mosaic in the greater area which includes densely wooded patches could be consistent (Vrba 1974, 1980, 1987b), which agrees with Reed’s (1997) conclusions.

I believe that our inferences of evolutionary process from paleoenvironmental reconstruction have been too crude and rigid. Consider three points in relation to the focus on the forest-to-savanna spectrum in the debate on hominid origin: (1) If a species, such as *Ardipithecus ramidus*, is found associated with “forest” and its inferred ancestor also lived in “forest,” there may have been differences in the nature of the forest which were not recorded in available evidence, but which had bearing on the evolution of the descendant species. (2) Even if the vegetation was identical, the descendant species may still have speciated by vicariance due to climatic or other physical change. Among living African mammals, there are many pairs of rainforest-adapted sister-species which speciated from forest-adapted ancestry following climatic vicariance (Grubb 1978). That is, the fact that the ancestral and descendant species both live(d) in forest does not mean that climatic change did not bring about speciation. (3) Even if the ancestral and descendant vegetational environments were identical, physical change may have altered other environmental aspects that brought new selection pressures and precipitated speciation. Patchiness does not only derive from vegetational heterogeneity. It can result from diverse factors including intrusive wetlands and other hydrological features, topographic barriers introduced by tectogenesis (such as rocky barriers), and lava or ash flows (e.g., Gulick [1872] showed that lava flows had very probably initiated speciation in the Hawaiian snail fauna). Selection pressure for traversing the barriers to reach the other side (which may have been what the *A. afarensis* individuals who formed the Laetoli footprints were doing; Leakey and Hay 1979) or for foraging in these areas (as in a shallow delta or wetland) might have been relevant to selection for onset or, later on, elaboration of bipedality. The notion that wading in shallow water played a part (Niemitz 2000; Verhaegen et al. 2002) seems reasonable given what we know about the paleoenvironments of many early hominid species.

There is some agreement that the onset of advanced bipedalism in *Homo* close to ca. 1.6 Ma not only falls during a time of change to more open and
seasonally arid landscapes but also makes sense as a selective response to these changes. Potts (1998) pointed out that by the latest Pliocene, populations of *Homo* were increasingly mobile, for example, tool-making behavior involved long-distance transport of stones as far as 10 km. Increased mobility is reflected by the migration out of Africa by 1.8 Ma of a lineage of *Homo* (if the early date for *H. erectus* in Java, Indonesia, is correct [Swisher et al. 1994]), the first of many subsequent migrations out of Africa which were associated with physical changes (Stringer 1995; Tattersall 1997b; Klein and Edgar 2002).

Other hominid phenotypes also appear at times of—and are adaptively consistent with—similar climatic changes. Stone tools appear by 2.6–2.5 Ma near the end of the large Late Pliocene cooling trend (de Heinzelin et al. 1999; Semaw et al. 1997). Hatley and Kappelman (1980) proposed that the climatic change led to this behavioral advance. They showed that a high below-ground plant biomass is characteristic of xeric open areas and argued that digging out of such foods, first by hand and later by digging sticks and other tools, evolved as an important feeding strategy of early hominids when the African savanna became more open and arid. From Bouri in Ethiopia comes the earliest evidence of tool use to butcher large mammal carcasses dated 2.5 Ma (de Heinzelin et al. 1999). The onset of more expanded tool kits appears to overlap with the climatic events ca. 1.8–1.6 Ma (Leakey 1971). Other behavioral changes may also date to the Plio-Pleistocene transition. For instance, the fact that the mandible and postcanine tooth crowns of *H. ergaster* (dated ca. 1.9–1.5 Ma), when scaled to body mass, are no larger than those of modern humans, may reflect earliest cooking (Wood and Richmond 2000). According to Wood (1995), the first signs of the “hypermasticatory trend” occurred with advent of *Paranthropus aethiopicus* ca. 2.6 Ma, followed by exaggeration in this trend ca. 2.3 Ma with the FAD of *P. boisei*, and further lesser modifications to the dentition of this species between 1.9 and 1.7 Ma.

The Late Pliocene and Pleistocene behavioral and cultural advances reflect reorganization and expansion of the brain. Although the available evidence indicates significant EQ increase in *Homo* only over the past 2 Myr (Holloway 1970, 1972, 1978; McHenry 1982), Holloway et al. (2003) presented evidence that brain reorganization predated brain expansion in hominid evolution. I previously suggested that the encephalization trend in *Homo* “evolved by progressive prolongation of ancestral, fast, early brain growth phases. It started with the modern ice age and was fuelled by progressive intensification of cooling minima since then” (Vrba 1996 p 15). I still suspect that we may find future indications that some of the brain modifications which came to characterize *Homo*—perhaps not increase in EQ but brain reorganization—were promoted by the start of the modern ice age. The largest EQ increase occurred between ca. 600–150 thousand year ago (opera cit.), toward the end of the mid-Pleistocene strong climatic events.
of ca. 1.0–0.6 Ma. Many selective scenarios for encephalization in Homo have been proposed. Falk’s (1980) review included warfare, language, tools and labor, hunting, and heat stress. Gabow (1977) emphasized population structure and culture, McHenry (1982) language, and Brain (2001) our predatory past. Vrba (1985a, 1988, 1989a) proposed that major selection pressures that led to brain and cultural evolution derived from the large-scale changes in climatic mean and amplitude during the Plio-Pleistocene; and that culture and the underlying brain modifications in Homo represent adaptation to eurybiomy or “generalist adaptation. Hominine culture is an extension of the common phenomenon in other animals that use behaviour to cope with climatic conditions. . . . a special case among animal behaviours that confers an expanded use of environmental resources” (Vrba 1989a p 30). As cited above, Potts (1998) made a similar proposal to explain the brain and behavioral adaptations of Homo. Others have also argued that Homo evolved toward biome generalization (Wood and Strait 2004).

The notion that robust australopiths, Paranthropus, were in certain senses specialists was originally proposed by Robinson (1963) based on the dentition of P. robustus. He suggested that the “crushing, grinding” robust vegetarian specialist lived in a somewhat wetter and more luxuriant environment than did the earlier gracile omnivore A. africanus. I suggested that the musculature of P. robustus was massive and the molars proportionally large “because their ‘vegetables’ were of the tough grassland type” (Vrba 1975 p 302) and that, in contrast to the more generalized Homo, “robust australopithecines were more specialized on open arid habitats” (Vrba 1989a p 30). The first really thorough analysis of the proposal that Paranthropus species were feeding specialists is that by Wood and Strait (2004). They concluded that Paranthropus species were most likely ecological generalists (i.e., eurybiomic in being able to make a living in varied environments) and made the novel proposal that (p 149) “. . . although the masticatory features of Paranthropus are most likely adaptations for consuming hard or gritty foods, they had the effect of broadening, not narrowing, the range of food items consumed.” I accept their arguments which accord well with available ecological information for Paranthropus. The acquisition, in response to newly encountered environments, of morphology which can perform a new specialized function, but which at the same time permits the retention of functions evolved in the ancestral more uniform environment, is a recurrent theme in the evolution of generalists. An example from other mammals occurs in the impala Aepyceros melampus which based on its cladistic placement (Vrba and Schaller 2000) evolved from a browsing ancestry. During the Plio-Pleistocene, this lineage evolved cranial and dental features which allow mastication of grass and other tough plant matter. The impala also has a stomach structure which
undergoes reversible seasonal changes (Hofmann 1973), a rare adaptation to varied vegetational environments. As a consequence of these dental and digestive evolutionary advances, the impala is today a consummate herbivore generalist which can subsist in different environments by switching its dietary intake.

A comparison of the hominid evolutionary changes with those in other mammals is beyond the present scope. But it is worth noting that there are numerous examples of anatomical and behavioral changes in mammals, which not only appear during roughly the same time periods when hominid novelties appeared but which also show similarities in the nature of the changes and in their ecological effects. Turner and Wood (1993) studied the times of appearance in the record of dental changes in various African mammals. They concluded that dental modifications for herbivory in more seasonally cold and arid conditions did show consistent temporal patterns among themselves and with those of some hominids (p 301): “changes in dental morphometrics support the interpretation of the development of savanna environments in response to colder and more arid conditions across the larger mammal fauna of eastern and southern Africa. The hypermasticatory development of Paranthropus was simply one facet of that response.” Many other mammalian changes accord with the climatic episodes summarized in Section 4.3.1 above. Examples include the increase in cursorial and migratory forms during major cooling trends (e.g., FADs during the start of the modern ice age of genera which today are migratory, such as Connochaetes in Alcelaphini [Vrba 1989a], Oryx among hippotragines [Vrba and Gatesy 1994], and the antilopine Antidorcas [Vrba 1995c]). Increased encephalization is linked to more advanced social organization and to life in more open, seasonally cooler and drier environments, not only in Homo but also in other mammals (e.g., see Oboussier’s [1979] results for EQ variation in a large range of living African bovids). An example of major modification of locomotory structure and function, convergently with what happened in hominids, is the evolution of bipedalism in rodents. As noted in Section 4.3.3, the new species of micromammals which first appeared during the cooling trend after 2.9 Ma included several bipedal steppe- and desert-adapted rodent taxa (Wesselman 1995) such as the first African and global appearances of the genus Jaculus of desert gerboas and a springhare species Pedetes. Hafner and Hafner (1988) pointed out that the bipedal forms share suites of characters—including enlarged hindfeet, heads, brains, eyes, and auditory bullae—while many of them also have larger body sizes. According to these authors, this highly specialized rodent body plan and bipedal locomotion is today strongly associated with open, arid, mostly desert habitats, and appeared numerous times independently in rodents (in 24 genera in 8 families). The world wide fossil record of bipedal rodents (Lavocat 1978; Savage and Russell 1983; Wesselman 1995) suggests that most may have appeared
during times of global cooling and land aridification either during the Late Miocene or the Late Pliocene.

4.4 Climate in relation to the evolution of ontogeny

4.4.1 Heterochrony pulses: parallel developmental responses to common environmental causes

The term heterochrony has been applied to both ecophenotypic and evolutionary changes in the rates and timing of ontogenetic events (Gould 1977). The same kind of heterochronic phenotype, $H$, commonly appears independently in different parts of a given monophyletic group in association with the same kind of environmental condition, $E$, and variously as an ecophenotype (i.e., reversible in later generations not faced by $E$) or as a phenotype, the expression of which is genetically fixed (or at least more constrained under varying conditions). An example is relative reduction of limb length in colder environments, an aspect of Allen’s Rule (Allen [1877]: mammalian extremities are reduced relative to body size in cooler climates). Not only the environmental association with $E$ but also the growth patterns tend to be similar between the independent occurrences of such a phenotype $H$ in a clade (Gould 1977a; Wake and Larson 1987; Vrba 1998b). It appears that certain kinds of heterochrony are more likely than others under particular environmental changes. Each heterochrony response starts off from the ancestral ontogenetic trajectory for that character, and this inheritance imparts limits and direction on what can grow and evolve. To the extent that aspects of ontogeny are shared by common inheritance between related species and across larger taxonomic groups, similar kinds of heterochrony will evolve independently in related lineages faced by the same environmental change. A summary and extension of the above is given in the following two statements (Vrba 2004, 2005):

1. Similar environmental changes elicit similar heterochronies in parallel, potentially in numerous lineages across large phylogenetic groups. Such heterochrony often involves change in body size and may be accompanied by large-scale phenotypic reorganization (Arnold et al. 1989; Vrba 1998b) such that the parallel heterochronies involve concerted evolution of suites of linked characters and “shuffling” among body proportions.

2. At times of widespread climatic change, diverse lineages may show parallel changes in size and in similar kinds of heterochrony associated in time and consistently with the climatic change—a “heterochrony pulse.” “Pulse” here
does not imply that the lineages responded in unison in a short time but only that the events are significantly concentrated in time.

I will mention one particular category of heterochrony, which is associated with body size increase by prolongation of growth and which is a common mammalian response to colder temperatures. It is of especial interest in the Plio-Pleistocene context of net global cooling; and it appears to have affected many African mammals including some evolutionary changes in *Homo*.

### 4.4.1.1 Cooling and body size increase

Many species with FADs during times of cooling and aridification were larger than their ancestral phenotypes (as cladistically inferred). For example, Vrba (2004) tested *H₀* that size changes across lineages are randomly distributed in time in the Alcelaphini (wildebeests, etc.) and Reduncini (waterbuck, etc.), which together comprise 63 recorded species over the past 5 Myr with a body weight range of ca. 20–250 kg. The result of significant peaks in size increase over 3.0–2.5 Ma and 1.0–0.5 Ma, two periods with strong cooling, is consistent with Bergmann’s Rule (1846: larger bodies are associated with colder temperature). While exceptions have been noted, in general the predictions are upheld in living mammals (Ashton et al. 2000; Meiri and Dayan 2003), including in humans (Baker 1988) and fossil mammals (Kurten 1959; Heintz and Garutt 1965; Davis 1981). To evaluate the claim that climate-associated heterochrony can involve extensive rearrangement—or “shuffling”—among body proportions, with parallel changes across related lineages, consider the example of Bergmann’s Rule. Bodies can become enlarged by faster growth relative to the plesiomorphic (or directly ancestral) ontogeny, by prolongation of growth time, or by a combination of both; and the influential factors may include temperature change itself or one of the attendant environmental changes (such as seasonal changes in food and water availability [Guthrie 1984; Barnosky 1986]). Such changes in growth mode are expected to result in rearrangement of body proportions. This is especially true of growth prolongation which is prevalent among Bergmann cases for which there are growth studies. For instance, many African tropical ungulates have shorter growth periods to smaller size in warm lowlands, while their close relatives at higher altitudes and/or latitudes grow for longer and become larger. The example of polymorphism in the African buffalo was noted earlier: *Syncerus caffer caffer* is much larger (up to 810 kg), grows for longer, and lives at higher latitudes and/or altitudes always near grassland, while the smaller and plesiomorphic phenotype *S. c. nanus* (up to 320 kg) with a shorter growth period lives in warmer, more forested regions.
Consider what is expected under the simplest way in which growth prolongation could occur: namely, if all ancestral growth phases for a character become proportionally prolonged (or extended in time by a constant factor) while maintaining the ancestral number of growth phases and the ancestral growth rates for respective phases (Vrba 1998b: Fig. 1). Let us call that simple proportional growth prolongation. Characters in the same organism have differing growth profiles in terms of growth timing and rate in relation to age and body weight (Falkner and Tanner 1986); and character growth typically occurs in distinct phases in each of which character change is nonlinear with respect to age (Koops 1986). We can distinguish two major types of heterochrony and associated allometric growth under growth prolongation: (1) In type A heterochrony, characters which grow with net negative allometry with respect to age and body size will become reduced relative to body size in the adult stage of the prolonged descendant ontogeny (even if no other growth parameter changes) and paedomorphic in that the descendant adult resembles the ancestral juvenile. A probable example is character evolution by Allen’s Rule (Vrba 1998b, 2004) which is upheld in modern humans (Baker 1988). The persistence of Allen’s Rule in modern biology supports the general hypothesis of similar changes in body proportions across lineages, which share inherited developmental responses to common environmental causes. (2) In type B heterochrony, characters which grow with net positive allometry become relatively enlarged. This mode, particularly by prolongation of a positively allometric late growth phase, may be how the hypermorphosed antlers of the giant Irish Elk evolved (Gould 1974) and how exaggerated secondary sexual characters in enlarged bodies commonly evolve (Vrba 1998b). As growth trajectories become prolonged, some characters become relatively reduced and others enlarged, with potentially extensive rearrangement among body proportions and substantial evolutionary novelty (Vrba 1998b: Fig. 1). Type B heterochrony can also result from prolongation of positively allometric early growth in which case the descendant structure is relatively enlarged and paedomorphic. An example is provided by the enlarged hindfeet of the bipedal, saltatory rodents during times of cooling (Section 4.3.4). If the growth of rodents, the juveniles of which in general have relatively large hindfeet (Hafner and Hafner 1988), is prolonged, a descendant adult with enlarged hindfeet is predicted. Evidence for at least some taxa is consistent with this, e.g., bipedal Kangaroo rats, Dipodomys, which inhabit semiarid to arid regions in North America, have longer growth periods and are hypermorphosed in some characters—yet paedomorphosed in others—relative to the ancestral ontogeny (Hafner and Hafner 1988). As noted earlier, the bipedal forms share suites of
characters in a characteristic body plan, which is today strongly associated with open, arid habitats and has appeared independently in 24 genera in 8 families (Hafner and Hafner 1988). I do not know how many of the 24 instances of parallel evolution involved growth prolongation. But I suggest that at least some of these appearances of suites of integrated character complexes exemplify coordinated morphological changes, by growth prolongation within and between lineages in response to a common climatic cause. This case illustrates that evolution by growth prolongation, as it acts on characters with different non-linear growth profiles in the same body plan, can result in a “shuffling” of body proportions. Substantial novelty in form can result, and also in function, as in these rodents which can jump to a height that is from 4 to 25 times their body length. I next discuss another example of type B heterochrony with prolongation of positively allometric early growth, namely encephalization.

### 4.4.1.3 Heterochrony and brain evolution

I applied statistical models for multiphasic growth to data on living human and common chimpanzee brain weights at ages since conception to test the hypothesis that encephalization of the human brain occurred by simple proportional growth prolongation (Vrba 1998b). Specifically, I wanted to know whether prolongation of the fetal growth phases, with strongly positive allometric growth, could account for most of the observed EQ increase. The results supported the hypothesis and imply that gross brain weight increase toward humans required change in only one growth parameter: prolongation of the nonlinear ancestral growth phases. In mammals in general, simple growth prolongation is predicted to result in encephalization as all mammalian brains complete a large proportion of their total growth rapidly early in ontogeny (Count 1947; Holt et al. 1975). The association in other mammals not only of body size increase with cooling but also of encephalization with more open, seasonally cooler, and drier environments, has already been noted. This raises the hypothesis that there were past “encephalization pulses” across many mammalian lineages in response to cooling over particular intervals (Vrba 1998b).

### 4.5 Conclusions

Environmental stimuli have influenced the evolution of hominids and other mammals at the levels of ontogeny, organismal adaptation, and speciation. The linkages to hominid adaptation have received most attention and some agreement has emerged in this area: successive cooling trends since the Late
Pliocene were associated with the earliest evidence of—and probably initiated—the “hypermasticatory trend” in *Paranthropus* (ca. 2.6 Ma) and its later exaggeration ca 2.3 Ma, stone tools and their use to butcher carcasses (ca. 2.6–2.5 Ma), Early Pleistocene expansion of tool kits, increased mobility by the Plio-Pleistocene interface and commitment to long-range bipedalism (ca. 1.6 Ma) in *Homo*, and significant brain expansion near 2 Ma and also since 600 thousand year ago. There is some consensus that encephalization and culture in *Homo* represent generalist adaptations which conferred a more flexible and expanded use of resources (Vrba 1989a; Potts 1998; Wood and Strait 2004). It now seems likely that the masticatory features of *Paranthropus*, while adaptations for consuming tough or gritty foods, had the effect of broadening, not narrowing, the range of food items consumed and allowed these forms to subsist in varied environments (Wood and Strait 2004). There is less agreement on environmental stimuli of the onset of bipedalism; in fact, doubt has been cast on the old “savanna hypothesis” by evidence ranging from locomotor anatomy of some early hominids to indications of dense tree cover in their environments. I discussed why, even if the hominid ancestor and its bipedal descendant species both live(d) in forest, this does not necessarily mean that climatic change did not bring about speciation.

Far less work has been done on the issue of environmental stimuli of hominid speciation. A brief summary of the current status is as follows: In terms of theory, the expectation that allopatric speciation predominates, particularly in hominids and other large mammals, is consistent with the weight of available evidence. It would take special pleading to argue that hominids are exceptions. If allopatric speciation predominates then so must physical initiation of speciation predominate. Most, and possibly all, of the hominid FADs either coincide with or fall very close to one of the major cooling trends. While taphonomic factors and chance may have contributed to this pattern, it does leave intact the hypothesis of climatic cause of hominid speciation. Also, on cladistic grounds, some speciation events must be closely associated with climatic change in hominids (Kimbel 1995) and other African mammals (Vrba 1995c). A study of the African larger mammal record of the past 10 Myr showed several turnover pulses which coincide with global cooling trends and that global cooling with increased aridity and seasonality was a more important stimulus of turnover than was global warming (Vrba 2000, 2005). A related finding is that since 10 Ma, the average large African mammal species started in a geographic distribution that was more restricted than later on (Vrba and DeGusta 2004), supporting hypotheses of speciation which emphasize the physical initiation of isolated populations.

Environmental stimuli of ontogenetic evolution have hardly been studied in our field. I discussed the “heterochrony pulse hypothesis”: the generative
properties shared among lineages can result not only in coherence of morphological changes but also in a strongly nonrandom timing of heterochrony events, as diverse lineages respond in parallel by similar kinds of heterochrony to the same environmental changes. This has not yet been tested. Of particular interest in the present Late Neogene climatic context is heterochrony involving body enlargement by prolongation of growth, because it is associated with colder (at least seasonally colder) temperatures (Bergmann’s Rule, upheld in modern humans [Baker 1988]). I have discussed some examples, including encephalization as a result of growth prolongation, in hominids and other mammals and have suggested that there were past “encephalization pulses,” across many mammalian lineages, in response to cooling trends over particular intervals such as during the onset and later intensification of the modern ice age.

In our field, one sometimes regrets (at least I do) that conclusive answers, such as emerge from some experiments of physical scientists, are so difficult to achieve. Hypotheses on the subject of environmental causes of hominid and other biotic evolution are difficult to test because the data come from different subdisciplines, each with its own set of biases and errors. As a result, debates tend to continue interminably. While we have a long way to go, on the positive side we can take heart in the simple fact that we are, so to speak, in a “growth industry”: while many aspects of life are deteriorating, the fossil record with its associated geological information is constantly improving. Thus there is a good expectation of future progress on some of the unresolved issues. In my view the results to date already offer support for the notion that common rules give qualitative and temporal coherence to the evolutionary responses across many mammalian—including hominid—lineages. These common rules arise from the regularities of physical change and from attributes of organismal ontogenies and phenotypes, and species, which are widely shared by common inheritance. The evidence implies closer linkages between the physical and biotic dynamics on earth than has traditionally been acknowledged. This perspective contrasts with a neo-Darwinian view: that selection of small-step random mutations is the vastly predominant evolutionary cause, with the implication that each evolutionary advance is to a larger extent an independent piece of history. Evolution is more rule bound than that, and our evolution is no exception.

References

Bergmann C (1846) Über die Verhältnisse der Wärmeeökonomie der Thiere zu ihrer Grösse. Göttingen Studien 31: 595–708
from the Upper Miocene of Chad, Central Africa. Nature 418: 801
Role of environmental stimuli in hominin origins


Feibel CS (1997) Debating the environmental factors in hominin evolution. GSA Today 7: 1–7


Gulick JT (1872) On the variation of species as related to their geographical distribution, illustrated by the Achatinellidae. J Nat 6: 222–224


Heintz A, Garutt VE (1965) Determination of the absolute age of the fossil remains of mammoth and wooly rhinoceros from the permafrost in Siberia by the help of radiocarbon (C14). Norsk Geol Tidsskr 45: 73–79


Pickford M (1991) What caused the first steps towards the evolution of walkie-talkie primates? Origine(s) de la bipédie chez les hominidés (Cahiers de Paléoanthropologie). Editions du CNRS, Paris


Tattersall I (1997b) Out of Africa again... and again? Sci Am 276: 46–53
Tattersall I (2000) Once we were not alone. Sci Am January: 56–62


Vrba ES (1987b) A revision of the Bovini (Bovidae) and a preliminary revised checklist of Bovidae from Makapansgat. Palaeontol Afr 26: 33–46


Vrba ES, Gould SJ (1986) The hierarchical expansion of sorting and selection: sorting and
selection cannot be equated. Paleobiology 12: 217–228


5 The Origins of Bipedal Locomotion

William E. H. Harcourt-Smith

Abstract
Bipedalism is a highly specialized and unusual form of primate locomotion that is found today only in modern humans. The majority of extinct taxa within the Hominini were bipedal, but the degree to which they were bipedal remains the subject of considerable debate. The significant discoveries of fossil hominin remains in the last 40 years have resulted in this debate becoming increasingly focused on how bipedal certain fossil taxa were rather than on the overall process. Although the early hominin fossil record remains poor, evidence points to at least two distinct adaptive shifts. First, there was a shift to habitual bipedalism, as typified by certain members of Australopithecus. Such taxa were bipedal, but also retained a number of significant adaptations to arboreal climbing. The second shift was to fully obligate bipedalism, and coincides with the emergence of the genus Homo. By the Early Pleistocene, certain members of Homo had acquired a postcranial skeleton indicating fully humanlike striding bipedalism. The final part of this chapter reviews why bipedalism was selected for. There have been many theoretical explanations, and the most robust remain those linked to the emergence of more open habitats. Such an environmental shift would have involved strong selection for new behavioral strategies most likely linked to the efficient procurement of food.

5.1 Introduction
Bipedal locomotion sets modern humans apart from all other living primates. We are the only obligate bipeds among well over 200 extant primate species. It therefore stands to reason that this unusual and highly derived form of locomotion has attracted much attention by those who study human evolution. Current evidence points to anatomical traits strongly associated with bipedalism relatively deep in the hominin lineage (Ward et al. 2001) and well before the advent of other “traditional” human traits such as larger brains and tool use. This chapter reviews the current state of thinking on this unique form of primate locomotion.
In order to understand the origins of hominin bipedalism, one first has to understand the mechanisms that make it such an efficient form of locomotion in modern humans. In the first section of this chapter, I will briefly explore the nature of the modern human walking cycle and the associated anatomical traits that facilitate it. I will then explore the fossil evidence for the origins of bipedalism and speculate on the likely locomotor behaviors that preceded it. Finally, I will discuss some of the theories surrounding why bipedal locomotion was selected for.

5.2 Locomotor differences between modern humans and great apes

Modern humans are fully obligate bipeds. After the first few years of life, it is the sole form of locomotion in all healthy individuals. By comparison, the great apes do not have any one form of specialized locomotion. *Pongo* is almost exclusively arboreal, but its locomotor behavior is taken up by clambering, vertical climbing, brachiation, terrestrial fist-walking, arboreal quadrupedalism, and even some above-branch assisted bipedalism. Orangutans are particularly well known to have a predilection for suspensory postures (Tuttle 1968). Clambering, which accounts for over 50% of observed locomotor behavior, mainly consists of forelimb suspension and hindlimb support and suspension (Tuttle 1968; Cant 1987). In this respect *Pongo* can be considered to be an arboreal specialist.

The most important aspect of the African apes is that, unlike *Pongo* and modern humans, their specialization lies not in their tendency to be either arboreal or terrestrial specialists but rather on having a mosaic of different locomotor modes that suit different environments and situations. Field observations have shown that all three taxa of African ape spend considerable time in both the trees and on the ground. The principal form of terrestrial locomotion is fast and slow knuckle-walking, where the legs do most of the propulsive work, and a significant degree of body weight is borne by the upper limbs through the knuckles (Tuttle 1970). African apes spend a small degree of time walking bipedally, but only for relatively short periods (Tuttle 1970; Hunt 1994). *Pan* also spends a proportion of its time standing bipedally, mainly to collect fruit in tall bushes, but it is important to note that even when doing so individuals are partially supporting themselves with their upper limbs, which are grasping onto branches (Hunt 1994; Doran and Hunt 1995). When in the trees, *Pan troglodytes* has a particular predilection for using knuckle-walking to move along large branches (Tuttle 1970).
5.2.1 The walking cycle

The modern human walking cycle is characterized by two distinct phases: the stance phase, when the leg is on the ground, and the swing phase, when it is off the ground. The stance phase begins with heel-strike, as the foot strikes the ground. The knee is fully extended and the foot dorsiflexed, which results in the heel striking the ground well before the rest of the foot. The foot then plantar flexes, and typically force is transmitted through to the substrate along its lateral border. The point when the body is directly over the weight-bearing foot is known as the midstance phase. The body then carries its forward momentum over the leg, at which point force moves medially over to the ball of the foot. At this point, strong muscular contraction of the plantar-flexors results in the ball of the foot pushing against the ground and eventually lifting away from it as the body continues to move forward. This action finishes with a final push-off of the big toe, known as toe-off. The leg is now off the ground and in the swing phase, with the knee and hip both bent so as to keep the leg off the ground as it swings forward to make the next heel-strike.

When chimpanzees walk bipedally there are considerable differences. The knees and hips remain bent throughout the stance phase, and the foot is less dorsiflexed at heel-strike. This results in a gait that is an awkward “shuffling” movement, with marked mediolateral swaying of the body from step to step, which is often referred to as a “bent-knee, bent-hip” (BKBH) gait. Heel-strike itself is at best weak and is often almost immediately followed by much of the rest of the foot making contact with the ground. There is little in the way of the lateral-to-medial shift in force transmission to the substrate during the late stance phase, and often three or more toes leave the ground at the same time (Elftman and Manter 1935).

5.2.2 Associated anatomical differences between humans and great apes

A large number of anatomical features are functionally related to bipedal locomotion, and it is the combination of these traits that allows this to be the sole form of locomotion in modern humans. Naturally, it is sometimes hard to determine which traits specifically facilitate bipedal locomotion and which are instead a result of it, but in terms of determining the locomotor affinities of fossil remains it is fair to assume that, either way, many of these traits certainly indicate bipedal locomotion to a lesser or greater degree. Table 5.1 summarizes the major anatomical features associated with human bipedal locomotion, but a
<table>
<thead>
<tr>
<th>Trait</th>
<th>Homo sapiens</th>
<th>African apes</th>
<th>Functional significance in <em>H. sapiens</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Foramen magnum orientation</td>
<td>Perpendicular to orbital plane</td>
<td>More vertically inclined</td>
<td>Related to vertical positioning of spine</td>
</tr>
<tr>
<td>Shape of spine</td>
<td>S curve with lumbar lordosis</td>
<td>C curve with no lumbar lordosis</td>
<td>More efficient balance and support of upright trunk</td>
</tr>
<tr>
<td>Intermembral index</td>
<td>Low (~72)</td>
<td>High (103–115)</td>
<td>Longer stride lengths</td>
</tr>
<tr>
<td>Size of vertebral bodies</td>
<td>Larger, especially L1–5</td>
<td>Smaller</td>
<td>Increased load of vertical trunk</td>
</tr>
<tr>
<td>Shape of iliac blades</td>
<td>Short, wide, and curved</td>
<td>Long, narrow, and flat</td>
<td>Support for vertical trunk</td>
</tr>
<tr>
<td>Orientation of iliac blades</td>
<td>Mediolaterally</td>
<td>Anteroposteriorly</td>
<td>Support for vertical trunk</td>
</tr>
<tr>
<td>Relative distance from hip to sacroiliac joints</td>
<td>Small</td>
<td>Large</td>
<td>More efficient transfer of weight from spine to hip</td>
</tr>
<tr>
<td>Size of acetabulum</td>
<td>Large</td>
<td>Small</td>
<td>Increase in weight-transfer through hip joint</td>
</tr>
<tr>
<td>Anterior inferior iliac spine</td>
<td>Present</td>
<td>Absent/Weak</td>
<td>Attachment site for strong iliofemoral ligament—helps maintain balance by preventing thigh from hyperextending</td>
</tr>
<tr>
<td>Femoral head size</td>
<td>Large</td>
<td>Small</td>
<td>Increase in weight-transfer through hip joint</td>
</tr>
<tr>
<td>Cortical bone distribution in femoral neck</td>
<td>Thicker inferiorly</td>
<td>Even all around</td>
<td>Increase in weight-transfer through hip joint</td>
</tr>
<tr>
<td>Bicondylar angle of femur</td>
<td>Valgus</td>
<td>Absent/Varus</td>
<td>Placement of lower leg closer to midline of body</td>
</tr>
<tr>
<td>Relative lengths of articular surfaces of femoral condyles</td>
<td>Medial condyle longer</td>
<td>Similar in length</td>
<td>Aids medial rotation of femur and locking of knee joint at heel-strike</td>
</tr>
<tr>
<td>Inclination of talar facet on distal tibia</td>
<td>Perpendicular to long axis of tibia</td>
<td>More laterally inclined</td>
<td>Allows perpendicular passage of leg over foot</td>
</tr>
<tr>
<td>Plantar tuberosity on calcaneus</td>
<td>Two</td>
<td>One</td>
<td>Facilitates stable heel-strike</td>
</tr>
<tr>
<td>Longitudinal arches in foot</td>
<td>Two present—one medial and one lateral</td>
<td>Absent</td>
<td>Acts as “shock absorber” and maintains structural rigidity in foot throughout stance phase</td>
</tr>
<tr>
<td>Hallux opposability</td>
<td>Absent</td>
<td>Present</td>
<td>Facilitation of efficient toe-off and loss of arboreal grasp</td>
</tr>
<tr>
<td>Relative tarsus length</td>
<td>Long</td>
<td>Short</td>
<td>Increases power arm length in foot—leads to more efficient leverage in foot</td>
</tr>
<tr>
<td>Relative lengths of rays II–V</td>
<td>Short</td>
<td>Long</td>
<td>Decreases level arm length in foot—leads to more efficient leverage in foot</td>
</tr>
<tr>
<td>Metatarsal robusticity pattern</td>
<td>$1 &gt; 5 &gt; 4 &gt; 3 &gt; 2$</td>
<td>$1 &gt; 3 &gt; 2 &gt; 4 &gt; 5$</td>
<td>Reflects increased transfer of weight along lateral edge of foot</td>
</tr>
<tr>
<td>Phalangeal curvature</td>
<td>Flat</td>
<td>Curved</td>
<td>Loss of arboreal grasp</td>
</tr>
</tbody>
</table>

Some descriptions adapted from Aiello and Dean (1990).
number of them warrant further discussion. In some cases, where particular traits relate to particular fossil specimens, there is further discussion in Section 5.3.

Some of the most radical morphological adaptations in the human skeleton that relate to bipedalism are found within the pelvis and lower limb. Compared to apes, the entire lower-limb complex in humans has become highly remodeled to cope with the intricate dynamics of balancing an upright trunk while efficiently moving the body forward. Balance is a particularly important factor as, at any point during the walking cycle, only one limb is actually in contact with the ground, and has to bear the entire weight of the body and balance it accordingly. The minimizing of mediolateral swaying of the body during walking is therefore critical, as it acts to stabilize the body over the supporting leg and reduce energy expenditure. Consequently, many of the traits associated with bipedal locomotion relate to two major factors: balancing the body as a whole, and keeping the downward transmission of force as close to the midline of the body as possible.

In modern humans, the foramen magnum is anteriorly situated and horizontally orientated (i.e., perpendicular to the orbital plane). This is a reflection of the vertical positioning of the spine. The spine itself has a distinct “S” shape, caused by marked lordosis in the lumbar region, which helps to bring the center of the trunk’s mass anteriorly (Fleagle 1999). The lower limb is considerably longer in modern humans than in great apes. *H. sapiens* has a low intermembral index (\(L/C24\), 72), whereas for *Pan* (103–106), *Gorilla* (115), and *Pongo* (139), it is far higher, reflecting their relatively shorter lower limbs and longer upper limbs. The longer lower limb in humans directly facilitates a longer stride length. The modern human pelvis is very different in shape to that of all other primates, including the great apes. The iliac blades are short and wide, the ischium extends posteriorly, and the sacrum is relatively wide. These features greatly facilitate support of the upright trunk, place the trunk’s center of gravity closer to the hip joint, and allow the lesser gluteal muscles to be positioned at the side of the pelvis (Napier 1967; Aiello and Dean 1990). This last feature is important, as contraction of these muscles during walking tilts the trunk toward the leg in contact with the ground, providing greater stability and balance. Humans also have a large acetabulum to accommodate a large femoral head, reflecting the relative increase in body weight passing through the hip during locomotion. The modern human femur has a valgus bicondylar angle, resulting in the knee being situated far closer to the midline of the body than the femoral head is. This greatly reduces the lateral deviation of body weight during walking, and it is argued by many to be an important feature related to habitual bipedal locomotion. However, modern humans who are unable to walk from birth do not develop a valgus bicondylar angle (Tardieu and Trinkaus 1994), and so this trait is best considered as epigenetic, even though its presence indicates habitual bipedal behavior.
The human knee has the unique ability to lock when in full extension, which greatly facilitates upright walking by keeping the leg straight and enabling the efficient downward passage of the body’s weight through to the ankle. This locking action is facilitated in humans by a relatively longer medial femoral condyle and different attachment sites of the femoral posterior cruciate ligament (Aiello and Dean 1990). The distal tibia has a particularly important feature linked to bipedal locomotion worth noting. The talar articular surface is orientated perpendicular to the long axis of the bone, resulting in a less arcuate passage of the leg over the foot (Latimer et al. 1987). This allows more efficient weight-transfer through to the foot. The modern human foot is particularly specialized for the requirements of bipedal locomotion. The African ape foot can be considered a grasping organ with some terrestrial adaptations, whereas that of humans is essentially a propulsive platform. Modern humans are the only living primates to have lost the ability to oppose the hallux, which is in line with the remaining toes. Human toes are relatively short and straight, and the tarsus relatively long, with an elongated calcaneal tuberosity. This allows for a more efficient lever-arm to power-arm ratio, which facilitates efficient propulsion during the stance phase. A combination of bony architecture and strong plantar ligaments results in the human foot being arched longitudinally on both the medial and lateral sides. By comparison, the ape foot is weight-bearing through the midfoot, and this is reflected in the enlarged medial tuberosity on the navicular. This longitudinal arching combined with the unique locking morphology of the calcaneo-cuboid joint allows the human foot to not only act as an efficient shock absorber, but also stay rigid during weight-transfer to the ground.

5.3 Fossil evidence

The precise number and nature of the derived traits characterizing stem hominins are difficult to determine and likely to remain so. It has recently been pointed out that, in its entirety, bipedality requires a combination of many complex anatomical traits and so cannot necessarily be classed as a dichotomous character (Haile-Selassie et al. 2004). However, it is reasonable to assume that strong evidence of bipedal locomotion is the key in determining whether fossil material warrants inclusion within the hominin clade. Perhaps the best way to consider such evidence is to ask whether fossil hominin material indicates habitual or obligate bipedalism, rather than merely the occasional bipedalism, which we see in most extant species of great apes (Rose 1991). In that context, it is clear evidence of a shift from occasional to habitual bipedalism, which is important when considering early hominin remains, and habitual to obligate bipedalism.
when considering relatively later hominin remains. Thus, when considering the hominin fossil record, this chapter will consider occasional bipeds to be those animals with a bipedal component of their locomotor repertoire similar to that of modern-day chimpanzees. By contrast, habitual bipeds are those taxa that had a significantly increased bipedal component but were by no means exclusively bipedal and would have retained an arboreal component to their locomotor repertoire. Obligate bipeds are considered as taxa that were exclusively bipedal and had lost all other forms of terrestrial and arboreal locomotor behaviors.

5.3.1 Precursors of bipedalism

There is a significant literature about the likely locomotor mode that directly preceded hominin bipedalism (see reviews by Richmond et al. 2001; Harcourt-Smith and Aiello 2004). In the absence of fossil evidence, early models relied heavily on observed extant primate locomotor behaviors and phylogenetic hypotheses. Arguably, the prevailing view was that a brachiating, hylobatid-like ancestor evolved into a larger-bodied African apelike ancestor capable of orthograde climbing and terrestrial knuckle-walking, which in turn evolved into a bipedal hominin (Keith 1903, 1923; Gregory 1916, 1928; Morton 1924, 1935). Minor variants of these models existed among authors, with Morton (1924, 1935) arguing for a more terrestrial “gorilloid” prehuman locomotor mode, while Gregory (1916, 1927) and Keith (1903, 1923) favored a more “troglodytian” hominin precursor. Others argued for a very deep tarsoid ancestry for humans and bipedalism (Wood Jones 1916, 1929) or for an arboreal quadruped ancestry of monkeylike above-branch locomotion (Straus 1949).

Despite the elegance of some of these early models, the central factor in understanding the evolution of bipedalism lies in the reconstruction of Late Miocene large hominoid locomotor behaviors. The advent of fossil evidence and molecular dating methods has effectively precluded some of these early theories from consideration. Based on molecular data, the last common ancestor of modern humans and chimpanzees is likely to have lived between 5 and 7 Ma (Gagneux and Varki 2000; Page and Goodman 2001), and most Miocene hominoid remains do not show a strong adaptation to brachiation (Napier and Davis 1959; Avis 1962; Rose 1991; Moyà-Solà and Köhler 1996), although the hylobatian model is still argued by some (Tuttle 1974, 1975, 1981). A number of alternatives have replaced these earlier theories in the recent literature. Perhaps the best-known recent theory is a suggested knuckle-walking ancestry for hominins (Washburn 1967; Richmond and Strait 2000; Richmond et al. 2001), which draws heavily on the specialized knuckle-walking behavior of chimpanzees and
gorillas as a model, and argues for a retention of traits associated with knuckle-walking in the wrists of *A. afarensis* and *A. anamensis* (Richmond and Strait 2000). However, this theory is disputed (Tuttle and Basmajian 1974; Dainton 1991; Lovejoy et al. 2001), and although certain Middle Miocene hominoid remains show an increased capacity for terrestriality, no large fossil hominoid taxa from the Middle or Late Miocene show adaptations for knuckle-walking behavior (Stringer and Andrews 2005). Alternative contemporary theories include those suggesting an arboreal climbing ancestor (either large-bodied or small-bodied) (Tuttle and Basmajian 1974; Stern 1975; Tuttle 1975, 1981; Prost 1980; Fleagle et al. 1981; Hunt 1996), a terrestrial quadruped ancestor (Gebo 1992, 1996; Sarmiento 1994, 1998), a *Pongo*-like pronograde-clambering ancestor (Crompton et al. 2003; Thorpe and Crompton 2005), and even an ancestor that practiced a type of terrestrial “tripedalism” with one limb always free to carry objects (Kelly 2001).

A universal theme that links both the older and the more recent hypotheses is the choice of a single specific locomotor mode as the dominant “precursor” to hominin bipedalism. As Rose (1991) points out, apart from humans, most primates usually use several different types of locomotor activity as part of their daily locomotor repertoire. Within the hominoid clade, *Pongo* and particularly *Hylabates* are traditionally considered rather derived and specialized in their locomotor behavior, while *Pan* and *Gorilla* are considered more generalized. However, the great specialization of *Pan* and *Gorilla* in fact lies in their particularly mosaic and versatile locomotor repertoire, especially in the case of the smaller-bodied *Pan*. Both genera regularly engage in terrestrial knuckle-walking, occasional bipedalism, vertical climbing, and orthograde clambering as part of their daily activities. It is quite possible that some of these behaviors, for instance knuckle-walking, may have been independently acquired in *Pan* and *Gorilla* (Begun 2004). Their locomotor behavior and associated anatomy, however, combined with our current knowledge of the Late Miocene fossil record, suggests that the immediate precursors to the very first hominins are likely to have been rather generalized hominoids (McHenry 2002) capable of a suite of different locomotor behaviors. In that context, it is perhaps rather limited to single out one particular locomotor mode as the likely “precursor” to habitual hominin bipedalism.

5.3.2 Evidence for habitual bipedalism outside the hominin clade

Although there is little current evidence to suggest that fossil hominoid taxa existing prior to the Hominini—Panini split had any significant degree of...
bipedalism in their locomotor repertoire, the possible locomotor affinities of one specific taxon are worth noting. There has been the suggestion that the Late Miocene European hominoid Oreopithecus bambolii was partially bipedal (Köhler and Moyà-Solà 1997; Rook et al. 1999). However, that assertion remains highly controversial. Although the iliac blades of Oreopithecus are reduced in length, it also has a suite of postcranial features that indicate adaptations to vertical climbing and forelimb suspension, including longer forelimbs than hindlimbs, a flexible shoulder joint, and a strong grasping foot (Harrison 1987, 1991).

5.3.3 Earliest hominin evidence

The earliest fossil evidence for potential hominin bipedalism comes from recently discovered Late Miocene cranial remains from Chad, dated to almost 7 Ma and assigned to the species Sahelanthropus tchadensis (Brunet et al. 2002). Virtual reconstruction of the distorted TM266 cranium is argued to show a foramen magnum that is more anteriorly positioned than in Pan and Gorilla and, more importantly, orientated almost perpendicular to the orbital plane (Zollikofer et al. 2005). This is a trait shared by modern humans and australopiths and indicates a more vertically orientated spinal column that is associated with bipedal locomotion. Currently, there are no known postcranial remains of S. tchadensis, precluding any further speculation on its locomotor behavior.

The femoral remains from the Lukeino formation in Kenya ascribed to a new putative hominin taxon Orrorin tugenensis (Senut et al. 2001) are also argued to indicate bipedal behavior. The material is dated to between 5.7 and 6 Ma (Pickford and Senut 2001; Sawada et al. 2002), and it is reported that there are anatomical features (specifically on the BAR 1002’00 proximal femur) that indicate habitual bipedal locomotion. The cortical bone of the inferior section of the femoral neck is argued to be relatively thick, and there is an “intertrochanteric groove” for the tendon of the obturator externus muscle on the posterior surface (Pickford et al. 2002; Galik et al. 2004). Thick cortical bone on the inferior section of the femoral neck is argued by some to imply habitual bipedalism (Pauwels 1980; Lovejoy 1988; Ohman et al. 1997), but others have noted that similar patterns of cortical distribution are found in many other primate species and that only apes and atelines differ in having relatively even distribution around the whole neck (Stern and Susman 1991; Rafferty 1998; Stern 2000). There is also debate as to whether the presence of the m. obturator externus tendon groove is reliable for inferring bipedalism. This feature, originally described by Day (1969), is argued to imply regular full extension of the thigh during bipedal locomotion.
Others argue that it is not a diagnostic trait of habitual bipedalism (Stern and Susman 1991) and that it can even be found in quadrupedal cercopithecoids (Bacon 1997). Most recently Lovejoy et al. (2002) claim that while the trait is completely absent in large samples of \textit{Pan} and \textit{Gorilla} and present in australopiths and 60\% of modern humans, it does not specifically imply bipedality but merely habitual extension of the femur. Finally, described upper limb morphology of \textit{O. tugenensis} includes a curved proximal manual phalanx and a humeral shaft with a straight lateral crest for \textit{m. brachioradialis} (Senut et al. 2001), both seen as adaptations for arboreal locomotor behavior (Senut 1981a, b, 1989; Stern and Susman 1983, 1991).

The only other hominin remains from the Late Miocene/Early Pliocene that possibly indicate bipedalism belong to the genus \textit{Ardipithecus} from the Middle Awash, Ethiopia (White et al. 1995). The 4.4–Ma \textit{Ardipithecus ramidus} was originally reported to have an anteriorly positioned foramen magnum (White et al. 1994), a trait that may indicate bipedal locomotor behavior (Lovejoy 1981). Older remains (5.6–5.8 Ma) recently ascribed to \textit{Ardipithecus kadabba} include one proximal fourth pedal phalanx that is described as having strong plantar curvature, but also a dorsally inclined proximal articular surface similar to that of \textit{A. afarensis} (Haile-Selassie et al. 2004). This latter trait is argued to show that \textit{A. afarensis} could dorsiflex its foot in a similar way to modern humans (Latimer and Lovejoy 1990b). However, it has also been argued that this feature in \textit{A. afarensis} is in fact intermediate between modern humans and great apes (Duncan et al. 1994).

In summary, the extremely meagre and fragmentary fossil record for early hominins precludes any serious speculation on how bipedal these taxa were. The horizontal orientation of the \textit{Sahelanthropus tchadensis} foramen magnum certainly indicates that this taxon was likely to have spent more time engaging in bipedal behaviors than either \textit{Pan} or \textit{Gorilla} do, but only the discovery of postcranial remains will further strengthen this argument. A combination of independent study and further postcranial finds are also needed to determine whether \textit{O. tugenensis} was anything more than an occasional biped. However, further postcranial remains of \textit{Ardipithecus} are currently under description (White 2002) and are likely to provide much needed insight into the locomotor repertoire of these early hominins.

### 5.3.4 The first habitual bipeds

Perhaps the first concrete evidence for habitual bipedalism comes with the earliest \textit{Australopithecus} remains from the Kanapoi and Allia Bay localities at Lake
Turkana, Kenya. Assigned to *A. anamensis*, the remains include a large and well-preserved distal and proximal tibia of one individual, and are dated to between 3.9 and 4.2 Ma (Leakey et al. 1995, 1998). Crucially, the distal end of the tibia has a horizontal talar surface relative to the long axis of the shaft, implying that the *A. anamensis* knee would have passed directly over the foot, as in later hominins and modern humans (Ward et al. 1999, 2001). In *Pan* and *Gorilla*, the talar surface is sharply inclined, which results in the knee passing over the foot more laterally during plantigrade locomotion (Latimer et al. 1987).

Following the *A. anamensis* remains, the record becomes far richer and starts with what is arguably one of the best-known and strongest lines of evidence: the Laetoli footprint trail. Laetoli, Tanzania, is the type locality for *A. afarensis* (see below) and has produced a number of hominin fossils assigned to this taxon. However, it is perhaps best known for its extraordinary series of preserved animal tracks, first discovered in 1976. Excavation through 1977 to 1979 revealed at least two (and probably three) trails of unmistakably bipedal hominin footprints preserved in a volcanic ash-fall layer that had become wet from rainfall (Leakey and Hay 1979; Leakey and Harris 1987; White and Suwa 1987). The footprints are dated between 3.5 and 3.7 Ma (Hay and Leakey 1982; Drake and Curtis 1987). The most distinctly hominid tracks are those from Site G, where there are two trails (and a possible third overprinted on the larger G-2 tracks). There is also a putative hominin track at Site A, although that has been argued by Tuttle et al. (1991) to have possibly belonged to an ursid.

Most researchers agree that the G-1 and G-2 series of tracks are very human-like, with no evidence of any type of forelimb support. The best-preserved prints show a strong heel-strike and toe-off and indicate a transmission of body weight through the stance-phase of walking similar to that of modern humans. In accordance with this, there is evidence of longitudinal arching, and the hallux is in line with the remaining toes (Day and Wickens 1980; Robbins 1987; Tuttle 1987; White and Suwa 1987). Stern and Susman (1983) argue that the footprints show a “transitional” morphology between apes and modern humans, but the prevailing view remains that they are very human-like. Most recently, Schmid (2004) has argued that although the prints were made by habitual bipeds, there is some evidence of increased rotational movement of the upper body reflecting a more ape-like morphology of the trunk. Schmid argues that this implies an ambling gait-pattern inconsistent with the ability to run.

There is much more debate over the taxonomic assignation of these trails. Most researchers accept that *A. afarensis* is likely to have made them, given that the type specimens for that taxon come from Laetoli and are roughly contemporary with the footprints. However, most noticeably, Tuttle and colleagues (1981, 1987, 1990, 1991) have argued that the prints are so human-like that they are...
5.3.4.1 Locomotion in A. afarensis and A. africanus

There has probably been more debate over the locomotor affinities of these members of the genus *Australopithecus* than over any other taxa. This is partially due to the fact that there is a relatively rich postcranial record for this genus. However, the main reason is that these remains show intriguing combinations of primitive and derived traits relating to both terrestrial and arboreal locomotor behaviors. Historically, the South African *A. africanus* remains provided the major focus of work through to the early 1970s, perhaps culminating in Robinson’s seminal treatise *Early Hominid Posture and Locomotion* in 1972. Between the 1970s and the 1990s, the discovery in Ethiopia of extensive postcranial remains assigned to *A. afarensis*, including the famous “Lucy” skeleton, has shifted the debate to East Africa, and back as far as 3.4 Ma. The *A. afarensis* remains are considerably older than those of *A. africanus* and along with Laetoli confirm that bipedal locomotion was likely to have been selected for well before brain expansion and tool-making behavior. In this section, I will review the morphology and associated locomotor behavior of *A. afarensis* and *A. africanus*, in turn.
**A. afarensis**  This taxon provides the first direct anatomical evidence of a true shift from occasional to habitual bipedalism. However, there has been considerable debate over the precise locomotor affinities of *A. afarensis*, most of which have fallen into two distinct camps. Some researchers argue that *A. afarensis* was almost as proficient a biped as modern humans (Latimer 1991; Lovejoy et al. 2002; Ward 2002). Others argue that in fact this taxon had a significant number of primitive postcranial traits that must have implied an important arboreal component to the locomotor repertoire (Susman et al. 1984; Stern 2000). Both these views are rather polarized, and it is best to consider *A. afarensis* as highly mosaic in its adaptations (see McHenry 1991, for a comprehensive review of primitive and derived traits in the Hadar hominins).

The first specimen of *A. afarensis* to be discovered showing evidence for bipedality was the AL 129 knee, discovered at Hadar, Ethiopia, and consisting of a well-preserved distal femur and associated proximal tibia (Taieb et al. 1974). Crucially, the morphology of the distal femur indicated a bicondylar angle even higher than that of modern humans (Johanson et al. 1976). This implied that the leg of *A. afarensis* would have fallen close to the midline of the body as in humans, which is an important adaptation to bipedal locomotion. Subsequent discoveries at Hadar, including the AL 288 partial skeleton (Lucy) and the extensive AL 333 assemblage, provided further evidence of a strong selection for bipedality. The AL 288 skeleton, approximately 40% complete, included a well-preserved pelvis, ribs, vertebrae, and representative pieces of all major limb elements. In overall morphology, the pelvis of Lucy is far more similar to that of modern humans. The iliac blades are short and wide, which would have allowed the lesser gluteal muscles to be situated laterally and act as pelvic abductors. The wide sacrum situated behind the hip joint would also have kept the center of mass of the trunk close to the hip, allowing efficient transfer of the weight to the lower limb during walking. Finally, there is a prominent anterior inferior iliac spine, indicating the importance of the knee extensor, rectus femoris, and a strong attachment of the iliofemoral ligament, which helps maintain balance by preventing hyperextension of the thigh (Aiello and Dean 1990). It is worth noting that the pelvis of AL 288 is also somewhat unique in being markedly wide, more so than humans, and that its iliac blades are not orientated as anteroposteriorly as they are in humans. This considerable width may well be functionally linked to the more funnel-shaped rib cage of *A. afarensis* (Schmid 1983, 1991). Such a rib cage would have been relatively wider inferiorly than in humans, therefore requiring a wider pelvis to support the resulting wider trunk. However, overall the features in the *A. afarensis* pelvis imply that it was well suited to two of the major requirements of bipedalism: maintaining balance and efficiently transferring weight from the trunk to the leg during walking.
Apart from the high bicondylar angle in *A. afarensis*, it has been argued that its long femoral neck is especially adapted to bipedality. This feature may in fact have made abduction of the hip biomechanically easier than in modern humans (Lovejoy 1973; Lovejoy et al. 1973). However, it is also possible that this feature is a reflection of the wider thorax and pelvis in *A. afarensis*. There are a number of other traits in the lower limb that unequivocally imply habitual bipedality. The mediolateral orientation of the talar surface of the distal tibia is horizontal relative to the long axis of the shaft. As discussed in Section 5.2.2, this is an important feature unique to later hominin bipeds that facilitates efficient transfer of weight from the leg to the foot. In the *A. afarensis* foot, the talus is very humanlike, particularly with respect to the trochlear surface (Latimer et al. 1987). The calcaneus also has a lateral plantar process on the tuberosity, which greatly helps diffuse stress as produced by ground reaction forces at heel-strike (Latimer and Lovejoy 1989). In general, the morphology of the ankle joint and heel in *A. afarensis* is extremely humanlike and would have been well suited to coping with the increased forces through the ankle associated with bipedal locomotion. It is also argued that the *A. afarensis* foot would not have been capable of opposing its hallux (Latimer and Lovejoy 1990a).

Other features present in the postcranium suggest bipedalism but are more open to interpretation. The femoral neck in both the AL 128-1 and Maka femora has thicker cortical bone inferiorly than superiorly (Lovejoy et al. 2002). As discussed earlier in Section 5.3.3, there is disagreement concerning the usefulness of this trait in inferring bipedality. The femoral condyles of larger-bodied members of *A. afarensis* are also more humanlike in proportions and symmetry, but this is not the case for smaller members of the species such as AL 129 (Aiello and Dean 1990). In the foot, it has been argued that *A. afarensis* had more dorsally orientated proximal articular facets on the proximal pedal phalanges, implying a humanlike ability for increased dorsiflexion of this joint in bipedal walking (Latimer and Lovejoy 1990b). However, a more recent metrical study has found that the *A. afarensis* angle actually falls well outside the human range of variation and between humans and the African apes (Duncan et al. 1994). The foot of *A. afarensis* had also been suggested to have had strong longitudinal arching (Latimer and Lovejoy 1989). This assertion is partially related to the assumption that the arched footprint from Laetoli were made by *A. afarensis*. As discussed earlier in Section 5.3.4, this may not be the case, and when assessing the degree of arching in this taxon it is best to assess the fossil remains directly. The markedly enlarged medial tuberosity on two navicular bones from Hadar strongly implies considerable weight-bearing in the midfoot of *A. afarensis* (Sarmiento 2000; Harcourt-Smith et al. 2002; Harcourt-Smith and Hilton 2005). Such morphology is incompatible with longitudinal arching. A recent
architectural analysis of *A. afarensis* pedal material by Berillon (2003) also finds that this taxon was unlikely to have had a longitudinal arch.

There are also a number of more ape-like traits in the *A. afarensis* postcranium, some of which suggest a degree of arboreal climbing ability (Stern 2000). Most noticeably, the manual and pedal proximal phalanges from the AL 288 and AL 333 localities are markedly curved and long and have prominent flexor ridges (Marzke 1983; Stern and Susman 1983; Susman et al. 1985). These features strongly imply an arboreal proficiency not found in later hominins. In the foot, the morphology of a partial medial cuneiform bone from AL 333 also implies that there may have been a degree of hallucial opposability (Harcourt-Smith et al. 2003), although others assert this to not be the case (Latimer and Lovejoy 1980a). Elsewhere, it has been reported that the morphology and function of the *A. afarensis* calcaneo-cuboid joint may have been apelike (Gomberg and Latimer 1984). Analysis of the limb proportions of the AL 288 skeleton show that the femur was relatively short (Jungers 1982; Jungers and Stern 1983), meaning that Lucy would have had a much shorter stride length than modern humans do. The morphology of the tibial plateau indicates that *A. afarensis* would have had a single attachment for the lateral meniscus, as in apes (Senut and Tardieu 1985), although the phylogenetic relevance of this trait has recently been questioned (Holliday and Dugan 2003). In the upper limb, it has been argued that the distal humerus of smaller-bodied *A. afarensis* specimens shows a well-developed lateral trochlear crest, an apelike trait that prevents dislocation of the elbow joint during climbing/suspension (Senut 1981a, b; Senut and Tardieu 1985), and a more cranially orientated glenoid (Stern and Susman 1983). Both these features could have facilitated above-branch climbing behavior. More recently it has been suggested that *A. afarensis* retained features in the wrist consistent with a knuckle-walking ancestry (Richmond and Strait 2000), although the authors do not go so far as to suggest that *A. afarensis* itself had a capacity for knuckle-walking. Others disagree with this assertion (Dainton 2001; Lovejoy et al. 2001), and it is interesting to note that none of the other important morphological traits associated with knuckle-walking (e.g., transverse dorsal ridges and dorsally expanded articular surfaces on the metacarpal heads) are found in *A. afarensis* specimens (Stern and Susman 1983).

It has also been suggested that there is a significant degree of postcranial variation between larger-bodied and smaller-bodied individuals of *A. afarensis*, particularly in the knee and elbow joints as discussed above, but also in the ankle (Stern and Susman 1983; Senut and Tardieu 1985). These differences could imply locomotor differences between the sexes, as Stern and Susman have suggested (1983), but have also been interpreted as suggesting that there were two distinct species of hominin at Hadar (Senut and Tardieu 1985; Tardieu 1985;
Deloison 1999). However, the prevailing view (Harcourt-Smith and Aiello 2004) remains that the Hadar material constitutes a single species, although there continues to be disagreement over the degree of sexual dimorphism in *A. afarensis* (Plavcan et al. 2005; contra Reno et al. 2003).

The considerable debate over the locomotor behavior of *A. afarensis* ultimately rests on how one views the relationship between these traits and the process of selection (Ward 2002; Harcourt-Smith and Aiello 2004). One can argue, as Latimer (1991) does, that the derived anatomical adaptations to bipedalism seen in *A. afarensis* demonstrate clear evidence of directional selection toward bipedality. Conversely, one can also argue that the retention of primitive apelike traits present in *A. afarensis* indicates a degree of stabilizing selection for arboreal proficiency (Stern and Susman 1983; Stern 2000), although there have also been suggestions that such features in *A. afarensis* were reflective of efficient terrestrial quadrupedalism (Sarmiento 1994, 1998). Only better understanding of the relationship between many of these traits and epigenetic factors may help to resolve this debate (Lovejoy et al. 2002; Harcourt-Smith and Aiello 2004). Overall, the postcranial skeleton of *A. afarensis* can be best considered as mosaic, showing a combination of derived human-like bipedal traits, primitive ape-like climbing-related traits, and a number of traits that appear to be unique. There is no doubt that *A. afarensis* was a habitual biped and would have spent a significant amount of time engaging in bipedal locomotor behaviors. However, there are also enough arboreal specialties present to imply a degree of climbing ability, and it is not unreasonable to posit that *A. afarensis* could have spent time in trees at night and for predator avoidance.

**A. africanus** Until the discovery of the Hadar remains in the 1970s, the South African *A. africanus* fossils provided the best insight into the locomotor behavior of ancient fossil hominins. This was initiated by the discovery of the Taung Child in the 1920s, and Dart’s (1925) description of its foramen magnum as being in a more humanlike position, thus implying upright posture and locomotion. Since then a large number of fossils assigned to *A. africanus* have been discovered, predominantly at the site of Sterkfontein. The specimen most diagnostic of bipedality is the partial skeleton Sts 14, which includes a parts of the pelvis, femur, and vertebral fragments, and has been recently argued to belong to the same individual as the Sts 5 skull (Thackeray et al. 2002). The pelvis is morphologically very similar to that of AL 288, in having wide and short iliac blades and being predominantly more humanlike than ape-like (McHenry 1986). Also like AL 288, the pelvis of *A. africanus* is very wide, with laterally flaring iliac blades, and has a relatively smaller acetabulum and iliosacral joint. This high-pelvic width is confirmed by other *A. africanus* pelvic fragments, including the recently
reconstructed Stw 431 pelvis from Member 4, Sterkfontein (Kibii and Clarke 2003). As for *A. afarensis*, this is argued to have provided a distinct advantage in bipedal walking (Lovejoy 1973). Distal femora from Sterkfontein (TM 1513 and Sts 34) also indicate that *A. africanus* had a high bicondylar angle, as in *A. afarensis* and modern humans. Recently discovered *Australopithecus* postcranial remains from the Jacovec Cavern at Sterkfontein, which may be as old as 4.0 Ma, include a proximal femur (Stw 598) that has a markedly long neck and small head, as for the *Paranthropus* femora from Swartkrans (see Section 5.3.4.2 below) (Partridge et al. 2003).

Other aspects of *A. africanus* locomotor anatomy have been argued to be more mosaic. McHenry and Berger (1998a, b) argue, mainly based on analysis of the Stw 431 skeleton, that *A. africanus* had relatively large upper limbs and small lower limbs, implying a more primitive climbing-related component of the locomotor repertoire. However, Stw 431 does not have any lower limb remains, only a partial pelvis with a preserved acetabulum and sacroiliac joint, making the scope for comparison limited. It is important to note that this study is often misinterpreted as stating that the limb proportions (e.g., humerofemoral index) of *A. africanus* were primitive. In fact, the study mainly concentrated on measurements taken from the articular surfaces at the ends of limb elements. Full recovery and analysis of the well-preserved Stw 573 skeleton from Sterkfontein Member 2 (Clarke 1998) will prove vital in helping to resolve this debate, as there are complete upper and lower-limb elements preserved. It has also been argued that a proximal tibia from Member 4, Stw 514a, is “chimpanzeelike” in having a more rounded lateral profile of the lateral condyle, thus inferring an apelike range of motion at the knee joint (Berger and Tobias 1996). This may have been so, but it is premature to describe a structure as complex as the proximal tibia as apelike based on one feature alone, and further analysis is needed. The discovery of the 3.3-Ma “Little Foot” partial skeleton (Stw 573) (Clarke and Tobias 1995; Clarke 1998) promises to one of the most important discoveries in the early hominin fossil record. The skeleton is far more complete than that of Lucy, with a complete skull, arm bones in articulation, leg bones, foot bones, ribs, and fragments of vertebrae and the pelvis (Clarke 1999, 2002). Most of these bones await removal from the breccia, but the foot bones were found separately and were initially described as showing a mosaic of adaptations, with a partially opposable hallux capable of some grasping potential but a more humanlike ankle joint (Clarke and Tobias 1995). However, recent metrical analyses of these remains shows that Little Foot could not oppose its hallux, had a navicular distinct from those at Hadar, but did have a more apelike ankle joint, implying that overall the foot was mosaic, but in a different way to that previously suggested (Harcourt-Smith et al. 2003; Harcourt-Smith and Aiello 2004). Finally, analysis of the relative size
of the semicircular canals in the inner ear indicates that \textit{A. africanus} had canals of apelike proportions. The morphology of the semicircular canals is closely linked to locomotor behavior, and while this finding does not preclude \textit{A. africanus} from having been a biped, it is likely that it would have been less competent at complex bipedal behaviors such as running and jumping (Spoor et al. 1994).

**Locomotor differences and similarities between \textit{A. africanus} and \textit{A. afarensis}** The postcrania of both \textit{A. afarensis} and \textit{A. africanus} show distinct adaptations for bipedal locomotion. Particularly for \textit{A. afarensis}, however, there is also strong evidence of retained apelike traits indicating a proficiency for arboreal climbing, especially within the upper limb. There is no doubt that both taxa were habitual bipeds, but at the same time they cannot be considered as obligate bipeds, and it is best to treat them as having had degrees of mosaicism in their locomotor repertoires. A number of studies have suggested that \textit{A. africanus} and \textit{A. afarensis} were very similar to each other in their locomotor anatomy (McHenry 1986; Dobson 2005). However, a number of other recent studies show that there is in fact a number of interesting differences between these two taxa. In a major analysis of the Stw 431 skeleton, Haeusler (2001) argues that a number of subtle but significant anatomical differences between the Stw 431 and Al 288 pelves to imply that \textit{A. africanus} may have had a different and more humanlike bipedalism. Recent work on the tarsal bones of australopithecines also shows that the putative Stw 573 \textit{A. africanus} foot may well have been mosaic (see above for details) in a different way to that of \textit{A. afarensis} (Harcourt-Smith 2002; Harcourt-Smith et al. 2003). Given that the Hadar remains and those from Sterkfontein (Member 2) could well be of an approximately similar geological age (Partridge et al. 1999), this suggests that there was perhaps a significant degree of locomotor diversity within \textit{Australopithecus}. \textit{A. afarensis} and \textit{A. africanus} both show significant adaptations for bipedality, but it is entirely possible that they achieved this through different evolutionary pathways (Harcourt-Smith and Aiello 2004).

### 5.3.4.2 Bipedalism in \textit{Paranthropus}

The majority of available postcranial material from the genus \textit{Paranthropus} come from the South African sites of Swartkrans and Kromdraai, and is assigned to \textit{Paranthropus robustus}. There are no complete long bones for \textit{P. robustus}, but from Swartkrans there are a partial pelvis (SK 50), two proximal femora (SK 82 and 97), and a number of other postcrania including hand and foot bones, while from Kromdraai there is a partial talus (TM 1517). Two major studies on this postcranial material by Napier (1964) and Robinson (1972, 1978) argued that \textit{P. robustus}
had a slightly less derived postcranial skeleton than *A. africanus* and would have had a less efficient type of bipedal gait. The main anatomical arguments for this were a more laterally facing acetabulum and longer ischium in the SK 50 pelvis, smaller femoral heads and a more medially orientated talar neck and head, which has been sometimes linked to hallux opposability (Broom and Schepers 1946; Napier 1964; Robinson 1972). However, the SK 50 pelvis is severely distorted, and it is questionable whether there is enough well-preserved morphology for serious anatomical analysis. The Kromdraai talus, although apelike in some metrical aspects (Wood 1974), also has a relatively flat humanlike trochlear surface (Robinson 1972), and the significance of talar neck orientation for grasping potential has been brought into question by Lewis (1980, 1989). More recent finds assigned to *P. robustus* suggest that it was likely to have been an efficient biped (Susman 1989). In particular, two well-preserved 1st metatarsals from Swartkrans show that *P. robustus* would have had a strong toe-off during walking, which accords with efficient bipedality (Susman and Brain 1988; Susman and de Ruiter 2004).

5.3.5 The rise of obligate bipedalism

In the Section 5.3.4.1 I discussed locomotor behavior within the genus *Australopithecus*. While there may well have been some diversity in the way that different species of *Australopithecus* were bipedal, what is certain is that they cannot be considered as fully obligate bipeds in the way that modern humans are. Conversely, later species of *Homo*, such as *H. erectus*, *H. antecessor*, and *H. neanderthalensis* were unequivocally obligate bipeds (see Figure 5.1 for a summary of which taxa there is agreement and disagreement over concerning bipedality). There are some subtle anatomical differences in the postcranial skeletons of these taxa when compared to modern humans, but their overall skeletal biology strongly implies fully humanlike bipedal locomotion (Trinkaus 1983; Aiello and Dean 1990; Lorenzo et al. 1999). It seems, then, that the emergence of true obligate bipedal locomotor behavior occurred between about 2.5 and 1.8 Ma. This time period is associated with the emergence of the genus *Homo* with which the emergence of obligate humanlike bipedalism is likely to be strongly associated.

This period is also extremely complex in terms of hominin evolution, and has been the subject of a diverse range of taxonomic interpretations. At least nine widely accepted hominin species have first or last appearances within this time frame, and the fossil record implies that there was considerable overlap in the temporal and geographical distribution of many of these taxa. Determining
Temporal ranges of known hominin taxa. Solid shading indicates taxa that were unequivocally obligate bipeds. Crosshatching indicates taxa where there is disagreement over the degree to which they were bipedal. No shading relates to those taxa where there is insufficient evidence. Adapted from Wood (2002)
which of these species were fully obligate bipeds and which were not has been
hampered by a number of factors. The principal issue is a meagre postcranial
fossil record, but even where there are significant numbers of postcranial ele-
ments, as at Koobi Fora (Leakey et al. 1978), there are often problems of reliable
taxonomic association. However, the 1.8‐Ma juvenile *Homo ergaster* skeleton
from Nariokotome, Kenya (KNM‐WT 15000) is shown to have been fully bipedal.
Its postcranial skeleton is remarkably humanlike, with long legs and short arms
and all of the derived postcranial traits associated with obligate bipedal locomo-
tion (Ruff and Walker 1993). With such an advanced body plan, it is reasonable to
assume that *H. ergaster* and possibly its direct precursors had developed obligate
bipedal behavior before 2 Ma. There are also a number of other postcranial
remains from Koobi Fora that imply striding bipedalism. In particular, the
temora KNM‐ER 1472 and 1481A are long and extremely humanlike. However,
it is difficult to speculate whether these specimens belonged to *H. rudolfensis,*
*H. ergaster,* or even to *P. boisei."

Most of the debate over the locomotor affinities of early members of *Homo*
has concentrated on postcranial remains assigned to *Homo habilis* from Olduvai
Gorge, Tanzania. Found at site FLK NN, the holotype for this taxon, OH
7, includes a number of predominately juvenile hand bones. These bones are
argued to be mosaic in their overall morphology. The scaphoid is apelike, the
proximal and intermediate phalanges are more curved than in modern humans,
and the intermediate phalanges have more apelike attachments for *m. flexor
digitorum superficialis,* a muscle associated with climbing and suspensory behav-
ior (Susman and Creel 1979; Aiello and Dean 1990). The OH 8 foot, also found at
FLK NN, is included as a paratype of *H. habilis,* and provides the best insight into
the locomotor behavior of this taxon (Day and Napier 1964; Leakey et al. 1964).
Extensive analyses of these bones indicate that the foot had strong longitudinal
arches, a locking calcaneo‐cuboid joint, a metatarsal robusticity pattern similar to
that of modern humans, and perhaps most importantly, a hallux in line with the
remaining toes that was wholly incapable of any opposability (Day and Napier
1964; Susman and Stern 1982; Berillon 1999, 2000; Harcourt‐Smith and Aiello
1999). The combination of all these features points to an individual capable of
efficient bipedal locomotion. However, the talus is less humanlike than the rest of
the foot and has a trochlea that is strongly grooved and medially sloping. This is a
more‐apelike morphology and is consistent more laterally arcuate passage of the
leg over the foot during the stance phase (Latimer et al. 1987). The implication of
this is that although the OH 8 foot is very humanlike in most critical features, its
ankle joint implies less efficient weight‐transfer from the leg during walking.
There are also the OH 35 distal tibia and fibula, which were found at site FLK
(Davis 1964). These are also argued to be humanlike, with a talar facet that is
perpendicular to the long axis of the shaft and predominantly human-like muscle attachments (Davis 1964; Lovejoy 1975; Susman and Stern 1982). It has been argued that OH 35 is likely to have come from the same individual as OH 8 based on morphological similarity (Susman and Stern 1982). Recent metrical comparisons contradict that assertion (Aiello et al. 1998), and indeed the two specimens were found 300 yards apart and in different geological horizons (Davis 1964; Dunsworth and Walker 2002).

The final specimen of interest is the more recently discovered partial skeleton OH 62, found at site FLK and assigned to *H. habilis* based on associated craniodental remains (Johanson et al. 1987). Although OH 62 is extremely fragmentary, it has been argued that its intermembral proportions were more apelike and similar to those of *A. afarensis*. This assertion was based on humero-femoral proportions that relied on an estimate of the femoral length of OH 62 as similar to that of the considerably older AL 288 (Johanson et al. 1987; Hartwig-Scherer and Martin 1991). Based on these findings, it has also been suggested that OH 62 has limb proportions as primitive as those of *A. africanus* (McHenry and Berger 1998a, b). However, the OH 62 femur is incomplete, lacking a considerable part the distal end, and it is impossible to accurately estimate the correct length of this fossil (Korey 1990; Haeusler and McHenry 2004). Furthermore, a recent reconstruction, based on morphological similarity to the younger (1.15–0.8 Ma) and undescribed OH 34 femur from Bed III, yields a far more human-like value, and therefore implies far more human-like limb proportions (Haeusler and McHenry 2004). Given that OH 34 could have been subjected to a degree of postdepositional erosion that may have compromised its morphology (Day and Molleson 1976), this latter finding must also be treated with caution. However, given that the OH 35 tibia and fibula are also relatively long, it is not unreasonable to assume that the limb proportions of *H. habilis* could have been rather more human-like than some have suggested. Until further evidence is uncovered, the evidence is not strong enough to definitively support either scenario.

There are a number of things that can and cannot be said about the locomotor affinities of *H. habilis*. The foot, tibia, and fibula are all very human-like in most critical aspects. There is some degree of uncertainty over whether the limb proportions were more human-like or apelike, but this issue cannot currently be resolved. The hand bones also show a mosaic of human-like and ape-like morphologies that may imply some climbing activity. Therefore, a conservative estimate of the locomotor behavior of *H. habilis* would place it between the habitual bipedalism of the australopiths and the obligate bipedalism of *H. ergaster* and later species of *Homo*. These findings imply a type of bipedalism in *H. habilis* more human-like and more efficient than those of either *A. afarensis* or *A. africanus*. 

The origins of bipedal locomotion
In summary, there is likely to have been some degree of locomotor diversity among different species of early Homo. The anatomy of the Homo ergaster postcranial skeleton (mainly based on KNM-WT 15000) is extremely human-like and derived, and it would have been an obligate biped capable of long distance travel (Wang et al. 2004). Recently it has also been suggested that this would have included endurance bipedal running, something that earlier hominins are unlikely to have been able to do (Bramble and Lieberman 2004). On the other hand, there is less certainty concerning Homo habilis. Although there is evidence of a distinct shift between the morphology and associated locomotor function of its postcranial remains and those of Australopithecus, it would not have been as efficient a biped as H. ergaster and is likely to have had a unique pattern of gait. Therefore, while it is certain that by the beginning of the Pleistocene fully obligate bipedalism had developed in some lineages of Homo, it cannot be argued that this had occurred in all species within that genus. This has been used by some to add weight to the argument that Homo habilis should be transferred to the genus Australopithecus (Wood and Collard 1999). This may or may not be the case craniodentally, but it cannot be argued postcranially. Further fossil discoveries will undoubtedly help resolve some of these issues. Recently, a number of new postcranial remains have been discovered at the 1.8 Ma site of Dmanisi in Georgia. Analysis of these remains could greatly aid our understanding of the emergence of the modern human body-plan in early Homo, as well as the nature of the locomotor anatomy of the first hominins to leave the African continent.

5.3.6 Summary of locomotor behaviors within the hominin clade

As we have seen, there are varying degrees of fossil evidence for the origins of bipedalism. More often than not, we are faced with the problem of being unable to place skeletal remains diagnostic of bipedality within a particular taxonomic hypodigm. Conversely, those fossil specimens which are associated with a particular species often show bewildering combinations of primitive, derived, and unique characteristics. However, on the basis of existing evidence a number of broad conclusions can be made (Figure 5.2). It is possible that the earliest hominins, such as A. ramidus, O. tugenensis, and S. tchadensis, show important enough features to imply a slight shift to increased terrestrial bipedality. However, current evidence is extremely meagre, and further finds may show that any shift to bipedality could have been more or even less substantial. Following these taxa, there appear to have been at least two distinct shifts in the development of hominin locomotion. First, between 4.5 and 3 Ma, a number of habitually bipedal
Figure 5.2
The degree of bipedality in known fossil hominins relative to time. Only taxa with documented traits relating to bipedal locomotion are included. *Ardipithecus, O. tugenensis,* and *S. tchadensis* are classed as occasional bipeds on the basis of having very few or weak traits related to bipedality. *Australopithecus afarensis* and *africanus* are classed as habitual bipeds on the basis of major anatomical remodeling of structures functionally related to bipedality, but retention of a number of apelike climbing specializations. *H. ergaster,* *H. erectus,* and later species of *Homo* are classed as obligate bipeds, but there is enough debate over *H. habilis* to place it between habitual and obligate bipedalism.
hominin species emerged, as typified by the Laetoli footprints and the extensive postcranial remains from Hadar. Within this time frame it is possible that different species varied in the way that they became bipedal, and that there were several different “types” of bipedalism being practiced. The period between 2.5 and 1.8 Myr heralds a second shift to fully obligate bipedalism. This period coincides with the emergence of the genus *Homo*, and by 1.8 Ma at least some members of this genus (i.e., *H. ergaster*) were fully obligate bipeds with a modern human body plan. Other early species of *Homo*, like *H. habilis*, may well have had a locomotor repertoire that was transitional between that of *Australopithecus* and that of *H. ergaster*. Subsequent to the advent of *H. ergaster* and *H. erectus*, all known hominins were fully obligate bipeds.

### 5.4 Why was bipedalism selected for?

As has been discussed, much contemporary debate over the origins of bipedalism rests on the locomotor adaptations of particular taxa or individual specimens. This is understandable given that fossils provide concrete evidence. However, it is critical to also ponder why bipedalism was selected for and why it became such a successful form of locomotion for our species. Most early theories as to why humans became bipedal center on the “freeing of the hands” as the principal force of selection. This can be traced back to Darwin, who argued in the *Descent of Man* (1871) that bipedal locomotion must have evolved to allow for the construction and use of hunting weapons. Since that first explanation there has been an abundance of different theoretical explanations, ranging from the plausible to the wholly implausible. When considering these different potential selection pressures, it is important to consider that bipedal locomotion is a highly derived and unique form of primate locomotion. In that context, we have seen that the skeletal modifications associated with bipedality are considerable. Strong selection pressures that specifically required prolonged periods of upright walking that are therefore likely to provide the key as to why bipedalism evolved (Lovejoy 1981; Rose 1991).

Prior to the discovery of the *A. afarensis* remains from Hadar, the orthodox view remained that tools and tool-use were intrinsically involved with the emergence of habitual bipedalism. Echoing Darwin (1871), some argued that tool use itself explained the selection for bipedalism (Washburn 1960), while others suggested that tool-carrying behavior was a more likely explanation (Bartholomew and Birdsell 1953; Washburn 1967). Both these theories are now contradicted by the temporal sequence of events provided by the contemporary fossil and archeological records (Rose 1991). Evidence of bipedal locomotion
currently predates the earliest stone tools by at least 1.5 Myr and probably more, which precludes the involvement of any stone-tool associated behavior in the origins of bipedality. More recent hypotheses have tended to be strongly linked to paleoenvironmental changes from the end of the Miocene through to the beginning of the Pleistocene. The emergence of habitual bipedalism strongly correlates with generally cooler and dryer global conditions, and an associated increase in more open grassland habitats (Van Couvering 2000). Predominantly forested environments were gradually replaced by more mosaic environments made up of different degrees of open grassland, bushland, and open woodland (Reed 1997). To cope with these environmental changes there is no doubt that hominins had to adapt a series of new behavioral strategies. Change in habitat composition would have resulted in a shift in food availability and thus necessitated a shift in food acquisition behaviors (Rose 1991; Foley and Elton 1998). Hominins either had to range further to find food or develop strategies to procure new and different types of food. It is therefore very likely that hominins would have had to have engaged in more terrestrial travel over more open habitats, and it is in this change in activity patterns that we are likely to find the reasons for the emergence of bipedality.

There are a number of more recent theories strongly associated with these selection pressures that warrant discussion. Lovejoy (1981) argues that food carrying and procurement by males was the driving selection pressure. This would tie in with some interpretations of the fossil material from Hadar that suggests that there was a degree of locomotor sexual dimorphism in A. afarensis (Stern and Susman 1983). Recent experimental work also supports Lovejoy’s (1980) theory in showing that introducing widely distributed “food piles” leads to an increase in chimpanzee and bonobo locomotor bipedality, mainly associated with food carrying (Videan and McGrew 2002). Such a situation could be analogous to the more spread-out concentrations of food sources available to hominins in a more open grassland environment. Increased bipedalism in such a setting would greatly increase the ability to carry food to desired locations. Other theories argue for terrestrial food gathering (Jolly 1970; Wrangham 1980), or even hunting (Carrier 1984; Shipman 1986; Sinclair et al. 1986). Jolly’s (1970) model uses the open-savannah geldada baboon as a modern-day analogue to suggest that early hominin bipedalism was linked to rapid seed-collecting behavior. Hunt (1990, 1994, 1996) has argued that chimpanzee postural behaviors may provide the key to our understanding of this issue. Over 80% of chimpanzee bipedalism is related to postural feeding. Using this as a behavioral analogue, Hunt argues that early hominin postcranial adaptations in Australopithecus were related to similar postural feeding behaviors and that true bipedal locomotion emerged with the advent of Homo (Wood 1993). It is
certainly possible that bipedal postural behavior may have preceded bipedal locomotion, but posture alone is likely to be too weak a selection pressure to have resulted in the significant anatomical remodeling seen in the *A. afarensis* and *A. africanus* pelvis and lower-limb structures (Lovejoy 1981; Rose 1991). It has also been recently suggested that bipedal threat displays may have been an important selective precursor to bipedal locomotion (Jablonski and Chaplin 1993).

One of the most interesting and widely accepted explanations of why hominins became bipedal is the thermoregulatory hypothesis suggested by Wheeler (1984, 1988, 1991, 1993, 1994). This argument rests on strong physiological explanations related to the reduction of thermal stress, and directly relates to the more open habitats that hominins would have been exposed to during the Early to Middle Pliocene. On the open savannah, quadrupedal animals expose considerably more of their body’s surface area to the sun. Wheeler calculated that a hominin standing upright would absorb 60% less heat at midday. Furthermore, being upright exposes the subject to any potential breeze, which would have a further cooling effect. These factors would greatly reduce the rate at which hominins would have overheated on open ground, meaning that they could have ranged further without having to have increased water intake. In a more open environment, where food sources were likely to have been more spread out, such an advantage would have greatly enhanced the ability of hominins to successfully collect food. One other physiological explanation for the development of bipedalism warrants comment. Rodman and McHenry (1980) have argued that there is a considerable energetic advantage to become bipedal. However, it has been shown that *Pan* and *Gorilla* locomotion is not any less efficient physiologically than that of modern humans (Steudel 1994).

Perhaps the most interesting point relating to all the above theories is that made by Robinson (1972), who states that there is unlikely to have been one specific reason why bipedalism was selected for. It was more likely a combination of several selective factors strongly relating to feeding strategies and reproductive behavior that provided the impetus for this shift in the hominin locomotor repertoire. Furthermore, bipedalism would have not only provided the ability to range further for food and other resources but also have exposed hominins to novel parts of the surrounding landscape, different types of predators, and new food sources. This in turn would have led to new hominin behavioral strategies to cope with such changes. It has also been argued that on the basis of increasingly variable environmental conditions during the Late Miocene and Pliocene, associated behavioral versatility would have been a critical selective factor for early hominins (Potts 1998). In this context, there is little doubt that selection for bipedality would have considerably facilitated such behavioral versatility.
5.5 Summary and concluding remarks

There is no doubt that the evolution of bipedalism is a critical issue in the study of human origins. However, as we have seen, there has often been a considerable degree of rather polarized debate and disagreement as to how, when, why, and in whom hominin bipedalism evolved. In particular, the emergence of so many important fossil finds in the last 40 years has resulted in the literature becoming increasingly “fossil driven” in its concentration on how bipedal a particular hominin taxon might have been. This has often clouded our understanding of the larger issues at stake surrounding the emergence of this unique form of primate locomotion. As Rose (1991) has pointed out, selection for bipedality was not an event but rather a series of processes. In that context, what can be said about these processes? It is certain that the selection pressures for bipedality must have been strongly linked to reproductive success, and it is therefore likely that such pressures would have been related to the efficient gathering and transport of food and other resources across increasingly open habitats. As discussed earlier in the chapter, the current fossil record points to at least one minor and two major steps in the emergence of obligate, humanlike bipedality. The earliest hominins were little more than occasional bipeds, while the australopiths can certainly be considered as habitual bipeds who still engaged in some arboreal locomotor behaviors. By the emergence of early Homo, certain species within that genus were unequivocally obligate bipeds much in the way that we are today. It is perhaps seductive to view such steps as punctuated events, and that may have been the case, but it is also possible that the fragmentary fossil record merely creates the illusion of such steps. Only the recovery and analysis of further fossil remains relating to bipedality, particularly from the Late Miocene hominoid record, will increase our understanding of this complex and unique process.

References

running and resource transport. Kluwer Academic/Plenum, New York


Keith A (1903) The extent to which the posterior segments of the body have been transmuted and suppressed in the evolution of man and allied primates. J Anat Physiol 37: 18–40


Reno PL, Meindl RS, McCollum MA, Lovejoy CO (2003) Sexual dimorphism in Australopithecus afarensis was similar to that of modern humans. Proc Natl Acad Sci 100: 9404–9409


Sarmiento EE (1994) Terrestrial traits in the hands and feet of gorillas. Am Mus Nov 3091: 1–56


running and resource transport. Kluwer Academic/Plenum, New York
Senut B (1989) Climbing as a crucial preadaptation to for human bipedalism. Ossa, 14: 35–44
Thorpe SK, Crompton RH (2005) The locomotor ecology of wild orangutans (Pongo pygmaeus abelii) in the Gunung Leuser ecosystem, Sumatra, Indonesia: a multivariate
Tuttle RH (1968) Propulsive and prehensile capabilities in the hands and feet of the great apes: a preliminary report. VIIIth Congress of Anthropological and Ethnological Sciences, Tokyo, pp 31–32
Wood Jones F (1929) Man’s place among the mammals. Edward Arnold, London
6 The Earliest Putative Hominids

Brigitte Senut

Abstract

Following the molecular studies, the hominid family has long been considered as emerging during the Pliocene. But today we have good evidence of hominids in the African Upper Miocene strata. However, reconstructing our earliest history is a difficult task as the Miocene data remain poorly known and fragmentary. Moreover, the tendency in anthropology to consider the modern chimpanzee as a good model for the last common ancestor of African apes and hominids has obscured our understanding of evolution because the apelike features are exclusively defined on the basis of a modern animal and not on those of Miocene hominoids. On the basis of detailed studies of Miocene apes and modern hominoids, it appears that bipedalism is probably the most reliable feature for defining hominids. Of the new hominoid taxa discovered in the Upper Miocene, only *Orrorin tugenensis* exhibits clear evidence of adaptation to bipedalism. At this stage, bipedalism in *Sahelanthropus tchadensis* and *Ardipithecus kadabba* still needs to be demonstrated. A common idea in hominoid evolution is that hominids emerged in dry, savannah-like environments; but the data gained from the Upper Miocene levels in Baringo (Kenya) clearly demonstrate that the environment of the earliest hominids was more forested and humid than expected. Finally, first discoveries of modern-looking apes have been made in 12.5Ma-old strata at Ngorora (Kenya) and 6-Ma old deposits at Kapsomin and Cheboit, indicating that the dichotomy between African apes and humans could be much older than generally thought.

6.1 Introduction

Identifying the earliest hominids remains a difficult task because the definition of the family varies widely from author to author. For some authors, the term hominid should be restricted to humans and their bipedal predecessors, whereas for others it should include all extant and fossil apes and humans; at its most extreme definition all African Great Apes should be included in the genus *Homo*.
(Czelusniak and Goodman 1998), or in a slightly less extreme view, only chimpanzees and bonobos (with the exclusion of Gorilla) should be grouped with Ardipithecus, Australopithecus, and humans in this genus (Wildman et al. 2003). In the latter scenarios, the search for the oldest hominid leads to a strange situation where the quest becomes that of identifying the earliest ape rather than the earliest humans. This is definitely not what the theme of the research on the origins of humans is today. The only consensus today among scientists (molecularists and anatomists) in ape evolution is that Pongo, Pan, and Gorilla do not belong to the same taxonomic group, a view that was widely accepted in the last century; the family Pongidae is now restricted to Pongo (Greenfield 1979). Homo is closely related to African apes and may be closer to Pan than to Gorilla. Whatever systematic scheme is considered, the focus is on understanding the ancestors of humans after their split from African apes. This is why it is more appropriate to restrict the term Hominidae to humans and their fossil forerunners. The widespread confusion in the use of the term “hominids” is due to the fact that authors usually do not define their position.

To understand the earliest hominids, we need to apply a geohistorical approach. The opposition between Asia and Africa in the history of our origins cannot be dismissed out of hand. However, today it seems clear from available field data that the development of the human lineage occurred in Africa.

Finally, we must consider the fact that, for the last 30 years, the history of research into the dichotomy between apes and humans has been dominated by the controversial results obtained by paleontologists on the one hand and molecularists on the other. The debate has focused on two major aspects: chronology and the search for the closest relative. There was a heated debate in the 1970s concerning the molecular clock and its implications for hominid evolution versus paleontological evidence and geological time. But discrepancies in the time scales between various neontological studies have never been thoroughly debated, as was pointed out by Arnason et al. (2001). We have known for almost two centuries that the African apes are our closest relatives, but the research published in the last three decades has attempted to focus on the question in greater detail and this has led to another major issue: is the common chimpanzee the closest relative to man? Or is it the bonobo? Or is it the group of African apes as a whole? At this point it has become widely accepted, almost without debate, that the closest relative to man is the chimpanzee, frequently claimed to share 98% or more of its genetic material with humans (or even 99.4% for some authors such as Wildman et al. 2003). This acceptance has occurred despite the fact that the problem is not yet definitively solved (Marks 2002). It is within this complex framework that research on our oldest ancestors has taken place during the past three decades. In addition, preconceived ideas about our earliest relatives make it
even more difficult to have a dispassionate discussion. A statement such as “the common ancestor of human and chimpanzees was probably chimpanzee-like, a knuckle-walker with small thin-enamelled teeth” (Pilbeam 1996) takes us back 200 years, being no different from the quest for the missing link. This widespread preconception is probably one of the reasons why most anthropologists have used modern chimpanzees as the basic comparative material when researching hominid origins.

This brings us to another aspect of the problem related to the definition of the earliest hominid. Most anthropologists consider that apelike (i.e., chimpanzee-like) features are primitive, and humanlike ones are derived. However, chimpanzees are not primitive; they are in fact highly derived in their locomotory and dietary adaptations and the use of their features as ancestral traits is a major error. Humans are also derived in their locomotion, but in a different way from chimpanzees. It is, therefore, not possible to define the polarity of these traits on the sole basis of some modern relict species; the Miocene apes were highly diverse and this diversity has to be considered when reconstructing phylogenies. The neontological approach, which has been in favor in some scientific circles, turns out to be a total failure when dealing with the definition of the earliest hominids. This approach leads to a search for magic traits, such as flat face, small canine, and thick-enamed teeth, which are considered almost universally to be hominid features: for when only modern chimpanzees and humans are compared, these features seem to be obvious and clear. However, it is necessary to understand their meaning and emergence before using them as a reference. When Miocene hominoids are included in the study, these same features are found to occur in many of them, suggesting that a flat face, small canines, and thick-enamed cheek teeth are plesiomorphic and that the elongated face of chimps, their large canines, and thin-enamed cheek teeth are apomorphies of the chimpanzee clade rather than plesiomorphies of hominoids. Exclusion of the fossils in the comparisons of modern apes and humans erases the diversity of the past which is the raw material for understanding our evolution.

6.2 The rise and fall of the Ramapithecus–Kenyapithecus group

During the 1960s and 1970s, it was widely claimed that Kenyapithecus (sometimes considered to be an African relative of the Asian Ramapithecus) was a hominid, aged ca. 15 Ma. This idea (Leakey 1961, 1962; Simons 1961; Simons and Pilbeam 1965; Andrews 1971) was questioned by some morphologists (Genet-Varcin 1969; Greenfield 1978). The divergence between apes and humans was thus considered
by these authors to be very ancient (about 16 Ma or even close to 20 Ma). But in
the 1970s, the development of molecular biology and the application of the
molecular clock led to the notion that the dichotomy was considerably more
recent, and ages of divergences for the African great apes and hominids of about
2–4 Ma were proposed (Wilson and Sarich 1969). It was in this context, 30 years
ago, that “Lucy” and the Afar australopithecines were discovered in 3 million-
plus-year-old deposits (Johanson and Taieb 1976). Considered at the time to
be the earliest hominid, these fossils were later formalized as *Australopithecus
afarensis* (Johanson et al. 1978). This major discovery was widely acclaimed: for
the first time, we could examine a quasi-complete fossil skeleton of an early time;
the earliest stages of bipedalism (a key feature in the definition of hominids)
could be seen, and we could get information on body proportions of the earliest
members of our family Hominidae. Not least, Lucy was considered by many to
occur in the range of dates estimated by the famous molecular clock, which at the
time suggested that the dichotomy between apes and humans took place about 4
Ma (Wilson and Sarich 1969). Subsequently, it was demonstrated that the
molecular clock was not reliable as it did not run at a constant rate in mammals
and in particular in primates (Britten 1986; Pickford 1987; Stanyon 1989). In
numerous papers, a large variety of dates was advanced for the dichotomy
between apes and humans, depending on the calibration ages accepted in
the studies and the type of protein used for the demonstration: these dates ranged
from 2.5 up to 4 Ma. But the fossils seemed always to give an earlier date. The
question of an early or a late divergence was addressed by Greenfield (1980):
he had already proposed that *Sivapithecus* (=*Ramapithecus*) *brevirostris* and
*Sivapithecus* (=*Kenyapithecus*) *africanus* were size variants of *Sivapithecus*, and
he suggested a late divergence for the group of *Homo*, *Pan*, and *Gorilla* from the
pongid stock. However, the age of this dichotomy that he published (10–5 Myr)
was greater than the one proposed by geneticists. Following general acceptance of
the molecular evidence that supported a divergence between chimpanzees and
humans around 4 Ma, the matter of chronology was thus thought to be solved
and neither *Ramapithecus* nor *Kenyapithecus* was subsequently considered by
many authors to be hominid. By the end of the 1970s, the *Kenyapithecus* material
had been restudied, especially in the light of sexual dimorphism in modern and
fossil apes (Greenfield 1978, 1979; Pickford 1986; Pickford and Chiarelli 1986). It
transpired that the group of Ramapithecines–Kenyapithecines did not belong to
Hominidae and as a result the Middle Miocene estimates for the dichotomy
between apes and humans were abandoned.

Eventually a meeting organized at the Vatican in 1982 led to a consensus
between paleontologists and molecularists: 7 Ma was the date of the divergence
adopted at this meeting.
6.3 Bipedalism and its impact on the origins of Hominidae

Among living primates, facultative bipedalism is frequent; but even if primates can sit with the back upright, stand on two feet, and walk bipedally for short distances, humans are the only ones that can move on two legs for long distances and extended periods of time. This difference is reflected in the skeletal characters of extant humans, often defined in comparison with chimpanzees. While a suite of features linked to femoral, pelvic, or sacral morphology appears to be soundly based, others are questionable. This is the case with the position of the foramen magnum. For the past 80 years, following Dart (1925), most scientists have considered that an anterior position of the foramen magnum indicates bipedality in hominids. Clark Le Gros (1950) used the anterior position of the occipital condyles to confirm the hominid nature of the australopithecines and proposed a “condylar position index,” but he noticed that in modern humans this position varied between dolichocephalous and brachycephalous individuals. Later (in 1972), the same author warned: “It has been assumed that the condylar-position index, by itself, is always correlated with the degree of postural erectness. The fallacy of this assumption is exposed by the fact that the index varies quite considerably even in modern H. sapiens.” However, generic, specific, and/or populational studies remained limited before 1960, despite the fact that Schultz (1955)...
highlighted the variability of this feature. Since then, it has been shown that an anterior position of the foramen is not linked exclusively with bipedalism but could be related to the development of the brain (Biegert 1963). Several authors demonstrated that its position relative to the cranial foramina was variable (Dean and Wood 1981, 1982; Schaeffer 1999; Granjean 2005). It is difficult to discriminate individuals on these isolated features as there is a great overlap between apes and humans.

Different forms of bipedalism have existed in the past. Of these, the most debated concerns Oreopithecus bambolii, discovered in Late Miocene lignites of Tuscany (Hürzeler 1958; Schultz 1960; Straus 1963; Sarmiento 1983; Tardieu 1983; Senut 1989) and which was recently demonstrated to be bipedal (Köhler and Moyà-Solà 1997; Rook et al. 1999). For the pedal features, these authors showed that this Miocene ape could move bipedally when on the ground, but with a stabilization morphology that differed from those of humans and australopithecines. Oreopithecus lived in an island environment where the absence of large predators and limited trophic resources played an important role in the evolution of mammals (Köhler and Moyà-Solà 1997).

6.4 The case of Australopithecus afarensis (=antiquus)

At one time every single fossil older than A. afarensis from Afar in Ethiopia was considered to be the earliest human ancestor; and this ancestor was almost always interpreted as being in the direct line leading to the genus Homo and then to us. However, this approach ignored or underestimated the probable diversity of Pliocene hominids. In fact, in the late 1970s, several authors had already pointed out that there might be a taxonomic problem with the species Australopithecus afarensis: was it, as claimed, a single bipedal species? Did this species include two different taxa, one of which was a combination of a climber and a terrestrial biped and the other a more advanced species which was primarily a ground dwelling biped? (see a review of Australopithecus afarensis locomotor adaptations in Stern 2000 and Coppens and Senut 1991). The difficulty derived mainly from the fact that before the 1970s, scientists had built their phylogenetic trees almost exclusively on the basis of dental anatomy, whereas the incorporation of locomotor traits led to a modification of these phylogenies. The picture became more complex in subsequent years, with a crop of new species of australopithecines being created; several of these had specimens also included in other hypodigms. This was especially clear with Praeanthropus africanus, Australopithecus afarensis, and Australopithecus anamensis. The use of cladistic methods did not clear up the problem (Strait et al. 1997; Strait and Grine 1999, among others), and
the phylogenetic approach became more and more confused, various scholars using the same species in different ways without defining them. 

As discoveries became more and more numerous, several genera were resurrected or created: *Praeanthropus*, *Ardipithecus*, *Orrorin*, *Sahelanthropus*, and *Kenyanthropus*. New specimens of a Pliocene hominid were found at Sterkfontein in South Africa (Clarke 1995). A major question remains today: is *Australopithecus afarensis* a direct ancestor or a side-branch of our family? This question is still unanswered.

### 6.5 The new challenge

The majority of scenarios concerning the dichotomy of apes and humans, with the exception of the East Side Story of Coppens (1983), failed to take into account the environment. Coppens’ hypothesis was eco-geographic in nature, the African Rift Valley constituting an ecological barrier between the apes in the west and early hominids in the east from about 7 to 8 Ma. But the most important elements of his hypothesis were chronological (the divergence took place between 8 and 7 Ma) and ecological (climatic change engendering modifications in regional vegetation patterns, etc.), whereas, despite its name, the geographic element was subsidiary in terms of the evolutionary scenario. As soon as we began to look for early and/or putative hominids in strata older than the Pliocene, we found them. In 2000, the discovery of early hominid remains (*O. tugenensis*) in the Upper Miocene strata of Kenya, and subsequent finds in Middle to Upper Miocene sediments of the same country, shed new light on the question of our divergence from the African apes. The *Orrorin* discovery was subsequently followed by finds in Ethiopia (*Ardipithecus kadabba*) and then Chad (*Sahelanthropus tchadensis*). The debate mainly focused on the C/P3 complex and on adaptations to bipedalism, and the status of these two species is still a matter of debate. But the main disagreement lies in the fact that in several studies comparisons were made basically with modern apes and later hominids, and very little with Miocene apes. As pointed out above, structures or features supposed to be hominid apomorphies might well be retained from older Miocene apes and some of the modern African ape features, usually considered to be primitive, might not be so.

#### 6.5.1 *Ardipithecus ramidus*

In 1994, *Australopithecus ramidus* was published and in 1995 it was attributed to a new genus *Ardipithecus*. This hominoid from Aramis localities 1–7 in the Middle
Awash (Afar Depression in Ethiopia) (White et al. 1994, 1995) was announced as the earliest known hominid. All the specimens, except the humerus (which was found above the Daam-Aatu Basaltic Tuff) come from a level located between the Daam-Aatu Basaltic Tuff and the Gaala Vitric Tuff complex. The tuff complex, situated at the base of the sections, has been dated at 4.39 ± 0.013 Myr and an age between 4.2 and 4.5 Ma can be estimated for the fossil hominid (WoldeGabriel et al. 1994). At the time of the discovery, these fossils were among the few supposed hominids older than 4 Ma. Recently, a few more specimens have been described from the Early Pliocene at As Duma in the Gona Western Margin (Ethiopia), the ages of which have been estimated at 4.51–4.32 Ma (Semaw et al. 2005). According to its discoverers, the new genus was based on differences from Australopithecus: the reduced megadontia of the postcanine teeth, the greater width of the upper and lower incisors compared with postcanine teeth, a narrow and obliquely elongated lower dm1 with a large protoconid and a small, distally placed metaconid without an anterior fovea, a small, low talonid with reduced cuspule development, absolutely and relatively thinner canine and molar enamel, lower P3 and upper P3 more strongly asymmetrical, with more dominant buccal cusps. With a canine that is not mesiodistally elongated, it is distinguishable from modern African apes. However, some of the cited features—including the thin enamel in the molars, asymmetrical upper and lower third molars, and the size relationships between the canines and jugal teeth—places Ardipithecus closer to the chimpanzee than to any of the oldest hominids known. The first deciduous molar shows resemblances to those of bonobos. But the morphology of the canine distances Ardipithecus from apes: it is more incisiform than in the latter group. Metric comparisons of the adult teeth were made with A. afarensis and underline the diminutive size of Ardipithecus. The upper canine/lower anterior premolar complex is typical of apes and was described as being “morphologically and functionally only slightly removed from the presumed ancestral ape condition” (White et al. 1994 p 308). However, certain features taken as support for its hominid status occur in female apes which have a reduced canine/premolar complex compared to those of males. Its postcranial bones reveal several apelike features, but the proximal humerus is more humanlike in the shallowness of the bicipital groove. However, this character occurs not only in hominids but also in other primates, such as Pongo, which is a cautious climber (Senut 1981). The fragment of occipital preserved would suggest that the foramen magnum is placed anteriorly relative to the carotid foramen, but for the reasons given above we must be cautious with the interpretation of this feature. At the end of 1994, a skeleton was found in the Aramis strata, but it has not yet been published. More recently, Semaw et al. (2005) briefly described a proximal third of a pedal proximal phalanx from the deposits of As Duma. They write: “The transversely
broad oval proximal facet is oriented dorsally, a character diagnostic of bipedality, and a trait also seen in *Ardipithecus kadabba*. But, Rose (1986) had already described the same feature in *Sivapithecus* from the Miocene of Pakistan. Given such weak evidence, it is difficult to accept bipedalism in *Ardipithecus ramidus*.

On the basis of the fauna, the botanical and sedimentological indications, the environment of *Ardipithecus ramidus* at Aramis is a forested one (WoldeGabriel et al. 1994). In the Gona sites, the faunal association, carbon isotopes, and sedimentology suggest a moderate rainfall woodland and woodland/grassland (Semaw et al. 2005).

### 6.5.2 Orrorin tugenensis

The discovery of *Orrorin* led to the elucidation of several aspects of early hominids (Senut et al. 2001). The specimens come from four sites: Cheboit, Kapsomin, Kapcheberek, and Aragai in the Lukeino Formation aged ca. 6 Ma (6–5.8 Ma) (Bishop and Chapman 1970; Bishop and Pickford 1975; Chapman and Brook 1978; Kingston et al. 1994; Pickford and Senut 2001) (Figure 6.1). The Lukeino Formation overlies the Kabarnet Trachyte dated by K/Ar, paleomagnetism, and biochronology at 6.1 Ma and is overlain by the Kaparaina Basalt, the age of which is estimated to be 5.7 Ma (Sawada et al. 2002). In the section, Cheboit and Aragai are the oldest sites, followed by Kapsomin and then Kapcheberek which lies in the upper level of the formation. Up to now, 20 specimens of *Orrorin* have been found consisting of the posterior part of a mandible in two pieces, a symphysis and several isolated teeth, as well as three femoral fragments, a partial humerus, a first phalanx, and a distal thumb phalanx. The genus is defined by its jugal teeth being smaller than those of Australopithecines, an upper canine short with a shallow and narrow vertical mesial groove and a low apical height, a small triangular upper M3, a lower P4 with offset roots and oblique crown, small *Homo*-like rectangular lower M2 and M3, thick enamel on the lower cheek teeth, a buccal notch well developed on the cheek teeth, no cingulum on the molars, a femur with a spherical head rotated anteriorly, the femoral neck elongated and oval in section, a medially salient lesser trochanter, a deep digital fossa, a humerus with a vertical brachioradialis crest, a curved proximal manual phalanx, and a dentition that is small relative to body size. *Orrorin* differs from *Australopithecus* in the morphology of the cheek teeth, which are smaller and less elongated mesiodistally. It differs from *Ardipithecus* by the greater thickness of enamel. It differs from both by the presence of a mesial groove on the upper canine. The upper and lower canines exhibit an apelike morphology, seen in female chimpanzees and Miocene apes; they are reduced in comparison with *Pan*. The apex of
the upper canine is pointed and almost sectorial, and a poorly developed lingual wear facet is visible.

The femora reveal that *Orrorin* was bipedal (Senut et al. 2001; Pickford et al. 2002; Galik et al. 2004). However, the other postcranial bones suggest that it could climb trees. The distal phalanx of the thumb exhibits features which are classically (but probably erroneously) associated with the manufacture of tools; these traits could be related to grasping abilities when climbing trees (Gommery and Senut, 2006).
At the time of its discovery, *Orrorin* was the first known bipedal hominid older than 5 Ma and indicated that the dichotomy between the African apes and the hominids had to be older than 6 Ma and that the classic recent dates of divergences estimated by molecular biologists did not fit with the paleontological evidence. On the other hand, the locomotor and dental features suggest that *Orrorin* was different from *Australopithecus afarensis*. It was a microdont animal with small postcanine teeth and a rather large body size, whereas *Australopithecus* was a megadont hominid with large postcanine teeth and small body size. Modern humans are microdont.

*Orrorin* did not live in an open environment but in a more forested one as suggested by the faunal remains, such as impalas, colobines, water chevrotain, arboreal civets, and floral remains, which contain large leaves with drip points (Pickford and Senut 2001; Senut and Pickford 2004; Senut 2006) from which it is concluded that in its early stages, bipedalism was not related to dry environments (Senut 2006). The humid conditions persisted in the Lower Pliocene (Pickford et al. 2004).

### 6.5.3 Ardipithecus kadabba

Material discovered in Ethiopia (Haile-Selassie 2001) from sediments aged between 5.2 and 5.7 Ma was identified as belonging to a subspecies of *Ardipithecus ramidus*, more recently raised to the specific rank *Ardipithecus kadabba* (Haile-Selassie et al. 2004). The material was collected at five different sites: Digiba Dora, Asa Koma, Alayla, Saitune Dora, and Amba East, from the Asa Koma Member of the Adu Asa Formation. The first four are in the Asa Koma Member of the Adu Asa Formation, and the deposits which have yielded the hominids are securely dated at 5.54–5.77 Myr by radiometric methods applied to underlying and overlying basalts. The Amba East material is slightly younger, being from the Kuseralee Member of the Sagantole Formation dated at 5.2–5.6 Ma (Renne et al. 1999). The morphology of the upper canine crown with a more rounded outline differs from *Orrorin*; but it also differs from *Ardipithecus ramidus* in the crest pattern of the same tooth, as well as in the morphology of the lower premolar which is more asymmetric in outline, and by the presence of a small anterior fovea. Moreover, the lingual cusps are more salient and sharp in the lower M3 and the upper M3 bears four cusps.

The species *Ardipithecus kadabba* differs from extant apes by its canines, which have a tendency to be incisiform as in *A. ramidus*, and by the presence of a clearly defined fovea on the lower P3, which is isolated from the mesial marginal ridge by a fold-like buccal segment. The postcranial morphology (Haile-Selassie 2001)
indicates several similarities to African apes and selected specimens from the Hadar, but the shape of the ulnar olecranon differs from that of hominids. A proximal pedal phalanx resembles the ones from Hadar, and on the basis of the dorsal orientation of the proximal facet of the bone, it supposedly belongs to a bipedal animal. However, the curvature seen in the *Ardipithecus kadabba* phalanx might be linked with arboreal adaptations as discussed by several authors (Stern and Susman 1983; Susman et al. 1984), and we must remain careful when assessing a locomotor complex on the basis of restricted material. *Ardipithecus kadabba* is associated with relatively wet and wooded environments as indicated by the fauna. As for the Amba East site, it seems slightly drier (WoldeGabriel et al. 2001).

### 6.5.4 *Sahelanthropus tchadensis*

The discovery of *Sahelanthropus* in Chad was published in 2002 (Brunet et al. 2002; Vignaud et al. 2002). Announced as the earliest known hominid, this status has been the subject of debate (Wood 2002; Wolpoff et al. 2002). It was found at Toros-Menalla (Chad) in deposits dated between 6 and 7 Ma, maybe closer to 7 Ma, by comparison with the Lukeino Formation and Nawata Formation (Vignaud et al. 2002). The following diagnostic features of the species have been published: orthognathic face and weak subnasal prognathism, small ape-sized braincase, long and narrow basicranium, small canines, robust supraorbital torus, absence of supratoral sulcus, marked postorbital constriction, small, posteriorly located sagittal crest and large nuchal crest, wide interorbital pillar, low-crowned jugal teeth and enamel thickness between that of *Pan* and *Australopithecus*, and anterior position of the foramen magnum. It is considered different from all the living great apes because of the relatively small canines, the apical wear of the canines, and a probable non-honing C/P3 complex. The claimed hominid status is based on the small, apically worn canine and on the structure of the face. However, when these complexes are considered among all fossil and extant hominoids, it appears that they are more frequent than believed. The maxillofacial complex in extant apes varies according to sex just as it does in Miocene hominoids (*Proconsul*, *Kenyapithecus*, *Ramapithecus*). It was this combination of features that originally led to *Ramapithecus* being proposed as a hominid, whereas it is today thought to be the female of *Sivapithecus*. Bipedalism in *Sahelanthropus* has been inferred from the position of the foramen magnum, but again, for the reasons expressed above, this feature can be misleading. The cranial base and nuchal area of *Sahelanthropus* (with its strongly developed nuchal crest and the flatness of the occipital) seem more apelike to some authors (Wolpoff et al. 2002),
suggesting a quadrupedal posture and locomotion despite the reconstruction proposed by Zollikofer et al. (2005), which fails to bring more evidence to the debate. The orientation of the plan of the foramen magnum falls within the range of variation of modern apes (Pickford 2005). Whatever Sahelanthropus is, its status as a hominid is still being debated (Wood 2002).

Sahelanthropus was found in perilacustrine sandstones and the sedimentological context suggests a mosaic of environments between lake and desert, which have been compared with the Okavango delta.

6.6 An earlier dichotomy?

During the past decade, several apelike fossils have been discovered in the Baringo District (Figure 6.2): a lower molar in the 12.5 Ma Ngorora Formation and three fragmentary teeth from the Lukeino Formation found in the same strata as *O. tugenensis*.

6.6.1 Ngorora

In 1999, a lower molar was collected at Kabarsero, Ngorora Formation, Tugen Hills (Pickford and Senut 2005) 12.5 Ma (Bishop and Pickford 1975). This tooth, probably a lower M2 (Pickford and Senut 2005), is close in morphology to *Dryopithecus* (Begun 2002) and chimpanzees and distant from similarly aged *Kenyapithecus* and *Otavipithecus* (Conroy et al. 1992; Ward and Duren 2002). An upper molar had already been described from Ngorora (Bishop and Chapman 1970; Leakey 1970; Bishop and Pickford 1975), but it appears to be closer in morphology to *Kenyapithecus* (Ishida and Pickford 1998; Pickford and Ishida 1998; Senut 1998), and the thickness of the enamel and the more centralized.
Figure 6.2 (continued)
cusps suggest that it belongs to a different hominoid from that of the lower molar. Two different kinds of hominoids would have coexisted at Kabarsero, a possibility which is also suggested in the Lower (Napak) and Middle (Moroto) Miocene sites of Uganda (Gommery et al. 1998, 2002).

If the derived characters of the Ngorora tooth are homologous to those of chimpanzees, then it would indicate that chimps were already a separate lineage by the end of the Middle Miocene, a suggestion that accords with some interpretations of the molecular data (Arnason et al. 2001). The resemblances between the Ngorora tooth and *Dryopithecus* indicate that the latter genus may have originated in Africa and migrated toward Europe about 12.5 Ma.

### 6.6.2 Kapsomin

In 2002, half an upper molar of a large hominoid was found at Kapsomin, Lukeino Formation, aged 5.9 Ma (Pickford and Senut 2005). This tooth is larger than those of *O. tugenensis* and the crown morphology is different. The trigon is wide, the distal fovea broad, the main cusps high and less inflated, and there is a deep buccal slit. The dentine penetrance is also high. Most of these features occur in *Gorilla* and are different from *Pan*.

In 2000, an upper central incisor was found in the same strata as *Orrorin*. Originally assigned to *O. tugenensis*, its morphology did not seem to fit with the early hominid. After a restudy of the specimen, it appears that it differs strongly from australopithecines and other hominids because of the lack of fossa on the lingual side of the tooth. Moreover, the crown is relatively low compared to root length, whereas in hominids the crown is higher with a scoop-shaped profile. In contrast, in *Gorilla* incisors, the lingual fossa is missing and the crown is wedge shaped.

In 2003, a lower molar was found at Cheboit, near the site of discovery of the first hominid tooth from the Lukeino Formation (Pickford 1975). The morphology of the tooth is compatible with the half upper molar from Kapsomin and the specimens probably belong to the same taxon. As for Ngorora and the Ugandan sites, two different hominoids would have coexisted at Kapsomin in the same strata, 6 Ma.

### 6.7 Conclusions

The debate about our earliest origins is probably not closed and is fueled by the poverty of fossils in the time period between 12 and 4 Ma. This is why it is
necessary to continue excavation and prospecting in different areas of Africa in order to fill the gaps and extend our knowledge of variation. One of the most troubling aspects of the research done to date on the origins of hominids is related to the comparative samples. Most scientists still focus on modern hominoids as good reference for primitive morphologies. However, these animals are highly derived in their cranial and postcranial anatomy. As long as Miocene apes are not properly considered in these studies, we will remain trapped in the quest for a mythical missing link.

Of all the features used to define hominids, probably the least controversial is bipedalism. We know that in the past there have been several types of bipedalism, but there is definitely a basic one that is known in australopithecines, *Orrorin* and *Homo*. In this group, adaptation to arboreality is variable: greater in some taxa, less in some others, and very little in *Homo*. There was probably a variety of early forms of hominids; the oldest widely accepted biped (supported by postcranial evidence) is *Orrorin* and we await further data on *Ardipithecus* and *Sahelanthropus* to clarify their status (*Figure 6.3*). What the evidence from the Upper Miocene tells us is that we cannot continue to support an origin of the earliest

*Figure 6.3*

Proposed relationships between early hominids and late apes
hominids in dry savannah-like conditions: in contrast, they inhabited humid to forested environments.

These early hominids cohabited with apes, and we are only just beginning to uncover the history of modern apes: this is the challenge of the third millennium.

References

Dean MC, Wood BA (1982) Basicranial anatomy of the Plio-Pleistocene hominids from
Pickford M (1975) Late Miocene sediments and fossils from the Northern Kenya Rift Valley. Nature 226: 918–921
Abstract

In this chapter, the historical, systematic, and anatomical evidence for the diversity of the species within the australopith grade is reviewed. Given a strict evolutionary species definition, nominal taxonomic diversity and species-lineage diversity do not necessarily map onto one another in the fossil record. Species–lineages entail statements of ancestry and descent that depend on the consistency of phylogenetic and stratophenetic data. The requirements for identifying species–lineages in the fossil record are severe, and in the early hominin record are rarely met, most often owing to small sample size, underrepresented character data, nonrepresentation of rare or short-lived taxa, poor chronological resolution, gaps in the time-stratigraphic framework, or some combination of these factors. Because hypotheses concerning the “bushiness” of the hominin phylogenetic tree depend on the identification of lineages, not phenteically based “paleospecies,” confidence with respect to this issue is not justified for the majority of the hominin fossil record. There are two cases in which an approach to this question can be attempted. In one, the evidence is consistent with the evolution of *Australopithecus anamensis* into *A. afarensis* via anagenesis. The other, the evolution of *A. boisei*, most likely entailed a speciation event that gave rise to southern African clade (represented by *A. robustus*) subsequent to the appearance of *A. aethiopicus*. The late Pliocene time period in which the latter events transpired (ca. 2.8–2.3 Ma) is one of substantial morphological diversity, high nominal taxonomic diversity, and high probability of synchronicity among known fossil samples. Although it is not possible to connect these australopith taxa (*A. aethiopicus*, *A. africanus*, *A. garhi*) to particular descendants due to defects in the data base, this time period probably documents a previously (and subsequently) unmatched degree of lineage proliferation compared to other parts of the human evolutionary record. The challenge to paleoanthropologists is to devote resources to improving this part of the fossil record and then to create testable phylogenetic and adaptive hypotheses to explain it.
7.1 Introduction

The australopiths constitute a taxonomically and adaptively diverse group of extinct hominins that are currently known to have inhabited the African continent between approximately 4.2 and 1.4 million years ago (Ma). The documented geographic distribution of this group includes present-day South Africa, Malawi, Tanzania, Kenya, Ethiopia, and Chad, but certainly may have been wider. Evidence from pelvic, knee, ankle and foot morphology, and where available trace fossils (i.e., the Laetoli footprint trails) indicates that these hominins were terrestrial striding bipeds, more similar, if not identical, in their locomotion to living humans (and extinct representatives of the genus *Homo*) than to any other known primate, living or extinct. The australopiths were impressively more variable in their skull and dental morphology than in their postcranial form, though the latter is poorly documented for some of the species included in the group, and our understanding of their taxonomic and adaptive diversity therefore reflects an historically important preoccupation with anatomical distinctions above the neck, particularly in the dentition and the bony structures thought to be associated with the masticatory system.

Many more characters distinguish the australopith species from one another than link them together as a group. Apart from bipedality, shared australopith characteristics include:

1. An approximately ape-size brain (range ca. 375–550 cm³)
2. Inferosuperiorly short, vertical mid-face with massive zygomaticomaxillary region and strong subnasal prognathism
3. Short basicranium with anteriorly positioned foramen magnum
4. Absolutely small, nonhoning canines
5. Large (in relation to body size) premolars and molars capped by variably thick enamel
6. Transversely thick mandibular bodies and tall ascending rami

Although these traits may, in combination, identify the australopiths, they do not diagnose them as a “natural” (i.e., monophyletic) group, as none of the listed features is unique to the group (i.e., they are not autapomorphies; they are either symplesiomorphies—shared with the great ape out-group to hominins, such as small brain size and strong subnasal prognathism, or synapomorphies—shared with one or more species in the closest sister taxon, *Homo*, such as bipedal locomotion, reduced canines, short cranial base, and, at least in comparison to some early species usually attributed to *Homo*, postcanine megadonty). As such, the australopiths most likely constitute a
paraphyletic group, identified by its unique adaptive grade among the hominins (small-brained, small-canined, megadont bipeds). This idea is reflected in the entrenched tendency to use the genus *Australopithecus* as a “waste basket” taxon for any and all of the extinct hominin species whose skeletal and inferred behavioral characteristics fit this adaptive pattern, notwithstanding the likelihood that one or more monophyletic groups, such as the so-called “robust” australopith species (sometimes attributed to genus *Paranthropus*), are thus subsumed within it.

### 7.2 Historical perspective on australopith diversity

The first described australopith species was *Australopithecus africanus* Dart 1925, identified on the basis of a fossil juvenile skull and brain endocast from the site of the Buxton Limeworks at Taung, Orange Free State (now Free State), South Africa. During the 1930s–1940s, no fewer than four additional species (in three genera) were identified based on adult craniodental fossils recovered from brecciated sediments within collapsed and eroded karstic structures then being mined for lime in South Africa (Sterkfontein, 1936; Kromdraai, 1938; Makapansgat, 1947; Swartkrans, 1948). The early taxonomy of the australopiths was authored by paleontologist Robert Broom (1950), who conducted the excavations at Sterkfontein, Kromdraai, and Swartkrans and perceived a different species at each site (Table 7.1). Broom’s protégé John Robinson subsequently collapsed this taxonomy to one species within each of two genera (Table 7.1) based on his scenario of a dichotomous partitioning of dietary resources by these early hominins (Robinson 1954). Robinson’s simplified taxonomy and hypothesis of a morphologically specialized herbivore *Paranthropus* and a more generalized omnivore (and presumptively meat-eating and toolmaking) *Australopithecus* has had a profound effect on subsequent australopith systematics. It remains

<table>
<thead>
<tr>
<th>Site</th>
<th>Broom (1950)</th>
<th>Robinson (1954)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taung</td>
<td><em>Australopithecus africanus</em> Dart 1925</td>
<td><em>Australopithecus africanus</em></td>
</tr>
<tr>
<td>Sterkfontein</td>
<td><em>Plesianthropus transvaalensis</em> (Broom 1936)*</td>
<td><em>Australopithecus africanus</em></td>
</tr>
<tr>
<td>Makapansgat</td>
<td><em>Australopithecus prometheus</em> Dart 1948</td>
<td><em>Australopithecus africanus</em></td>
</tr>
<tr>
<td>Kromdraai</td>
<td><em>Paranthropus robustus</em> Broom 1938</td>
<td><em>Paranthropus robustus</em></td>
</tr>
<tr>
<td>Swartkrans</td>
<td><em>Paranthropus crassidens</em> Broom 1949</td>
<td><em>Paranthropus robustus</em></td>
</tr>
</tbody>
</table>

* Broom initially assigned the species *transvaalensis* to *Australopithecus*. 
the leading explanatory paradigm for the evolution of these hominins to the present day.

Much of the early debate about the australopiths’ role in human evolution had less to do with their adaptations or diversity than with their suitability as human ancestors. The skepticism that greeted Dart’s claims of human ancestral status for *A. africanus* is well known, and despite the subsequent demonstration by Broom (1939; Broom and Schepers 1946), Gregory and Hellman (1939); Clark (1947) and others of uniquely human characteristics in adult australopith teeth, jaws, and basicrania, it was not until the identification of humanlike bipedal adaptations—and the absence of apelike quadrupedal ones—in a distal femur (TM 1513) and capitate (TM 1526) from Sterkfontein and a talus (TM 1517) from Kromadraai that the most ardent skeptics capitulated (e.g., Keith [1947], though his recantation began earlier: from Robert Broom [Broom and Schepers 1946 p 22]) we learn that “In a letter I have recently received dated 11th May 1944, Sir Arthur writes: ‘No doubt the South African anthropoids are much more human than I had originally supposed, and I am prepared to swallow plantigrade adaptations in their limb bones.’”

Although australopith fossils had lurked unidentified in east African faunal collections as early as 1935 (i.e., the BMNH M 18773 lower canine from Laetoli, Tanzania [White 1981], it was not until the discovery of the Garusi (Laetoli) maxillary fragment in 1939 that attention was directed (albeit fleetingly) toward the East African Rift Valley as a source of early hominin fossils. The particularly apelike upper premolar and palate morphologies of the Garusi specimen were overshadowed by the powerful appeal of Robinson’s adaptive scheme in which the fossil was cast as an east African representative of *A. africanus* (Robinson 1953a), and thus the Garusi maxilla receded into relative obscurity until the recognition of the craniodentally plesiomorphic *A. afarensis*, based on new mid-Pliocene Hadar (Ethiopia) and expanded Laetoli collections in the late 1970s (Johanson et al. 1978; see below).

Mary Leakey’s 1958 discovery of the megadont *Zinjanthropus boisei* cranium OH 5 at Olduvai Gorge, Tanzania, together with discoveries in the 1960s–1970s at Peninj, near Lake Natron in Tanzania, in the Omo River basin of southern Ethiopia, and at Koobi Fora, east of Lake Turkana, Kenya, populated the Plio-Pleistocene fossil record of eastern Africa with crania, jaws, and teeth of a hominin whose morphological pattern was easily accommodated in the adaptive mold of Robinson’s vegetarian *Paranthropus* (Robinson 1960). Most workers have since upheld separate species status for these hominins (as *P. boisei* or *A. boisei*) on the grounds of morphological distinction and/or geographical separation (Tobias 1967; Howell 1978; Rak 1983).

The identification in 1964 of *Homo habilis* approximately contemporaneous with OH 5 at Olduvai Gorge (in Bed I, ca. 1.7 Ma) recalled the earlier discovery of
“true man” (*Telanthropus capensis*) alongside *Paranthropus* in the Pink Breccia (later Member 1) at Swartkrans (Broom and Robinson 1949; Robinson 1953b; Leakey et al. 1964). Robinson (1965 *et seq.*) dismissed the Bed I *H. habilis* fossils as an east African variant of *A. africanus*, which he believed was ancestral to modern humans, while advocates of the “single species hypothesis” struggled to accommodate all southern and eastern African hominin morphological diversity within single evolving species (Brace 1967; Wolpoff 1970). However, the recovery of a cranium of *Homo erectus* (KNM-ER 3733) in the same radioisotopically dated horizon (ca. 1.75 Ma) of the Koobi Fora Formation (Kenya) that yielded *A. boisei* (e.g., KNM-ER 406) validated once and for all the idea that the pattern of early human evolution was, at least partly, the result of a process of diversification through speciation rather than a linear process of morphological advancement (Leakey and Walker 1976).

The Leakeys’ argument for a geologically early appearance of genus *Homo*, bolstered by the 1972 recovery of the relatively large-brained, flat-faced KNM-ER 1470 (initially dated erroneously to >2.6 Ma but now known to be ca. 1.9-myr old), strongly influenced the first interpretations of fossil hominin diversity in the collections made during the 1970s in pre-3.0-myr-old sediments at Hadar and Laetoli. While the small Laetoli sample was thought to contain an early precursor of *Homo* (Leakey et al. 1976), as many as three hominin taxa, including one similar to the Laetoli form, were believed to inhabit the Hadar assemblage of teeth, jaws, and postcrania (Johanson and Taieb 1976). Such early diversity made sense given the Robinsonian view of an *A. africanus*-like ancestor for a geologically old *Homo* lineage, implying earliest Pliocene or even late Miocene hominin cladogenesis (Tobias 1973).

By the late 1970s, the correct allocation of KNM-ER 1470 to the latest Pliocene and the reinterpretation of the Hadar and Laetoli fossils as representing a single species (Johanson and White 1979) had refocused discussion around two different two-lineage models, both of which featured *A. afarensis* at the root of mid-Pliocene hominin diversity. One, a twist on Robinson’s diet-driven diversity scenario, envisioned *A. africanus* as the basal taxon of a “robust” australopith clade, based on the view from the perspective of the craniodentally plesiomorphic *A. afarensis* of derived masticatory morphology in the hominins from Sterkfontein, Makapansgat, and Taung (White et al. 1981; Rak 1983; Kimbel et al. 1984). The other interpreted *A. africanus* as the better candidate for the common ancestry of *Homo* and “robust” australopiths by virtue of their shared derived morphology in the premolars and canines, mandible, and calvaria relative to more plesiomorphic states in *A. afarensis* (Kimbel et al. 1984; Skelton et al. 1986), but with implied reversals in derived states inferred to be related to heavy chewing in the transition to *Homo*. 
Underscoring an emerging consensus on the importance of cladogenesis in the early record of human evolution, homoplasy became a dominant theme in paleoanthropological writings on australopith systematics during the 1980s. These focused in particular on the likelihood of independent evolution of skeletal responses to heavy mastication (see contributions in Grine 1988), an issue highlighted by the discovery of the 2.5‐myr‐old KNM‐WT 17000 cranium attributed to *A. aethiopicus* (Walker et al. 1986) whose startlingly plesiomorphic attributes within an otherwise fairly typical “robust” australopith anatomical milieu ensured a high degree of homoplasy (in the masticatory apparatus or cranial base) in any phylogenetic hypothesis that attempted to accommodate all australopith (and early *Homo*) species then known (Kimbel et al. 1988). *A. africanus*, more than any other species, continued to occupy a place of uncertainty in hominin phylogeny, which led to musings on whether the type species of the genus may actually comprise more than one species taxon (Clarke 1988, 1994; Kimbel and White 1988; Kimbel and Rak 1993)—a view that remains in the minority despite the uncommon degree and type of cranial and dental morphological variation within the now 700+ specimen Sterkfontein sample.

The 1990s was a decade in which a great deal of paleoanthropological field work concentrated on extending the record of early hominins into the late Miocene, the period in which the evidence from DNA pinpoints the divergence of chimpanzee and human lineages. At least five hominin taxa predating *A. afarensis* have been recognized since 1994 (*Ardipthecus kaddaba*, *Ard. ramidus*, *Orrorin tugenensis*, *Sahelanthropus tchadensis*, *A. anamensis*), aggregately spanning the period between ca. 4.2 and ca. 6–7 Ma. Relative to *A. afarensis*, all are more plesiomorphic in dental, mandibular, and cranial morphology. The youngest, *A. anamensis*, is by far the most similar morphologically to *A. afarensis* and subsequent australopiths (Leakey et al. 1995; Ward et al. 2001). For the purposes of this chapter, the australopiths will be taken to include *A. anamensis* as their earliest known representative.

Other recently proposed taxa in the 2- to 4‐myr time range, including *A. garhi* from Ethiopia (Asfaw et al. 1999), *A. bahrelghazali* from Chad (Brunet et al. 1996), *Kenyanthropus platyops* from Kenya (Leakey et al. 2001), and a so far undiagnosed form from Member 2 at Sterkfontein, South Africa (Clarke 2002), hint at as yet poorly understood aspects of hominin diversity and geographical distribution in the early to middle Pliocene of Africa. They will be mentioned in the survey of australopith species provided below, but relatively little is known about them compared to the other species included in the group. Table 7.2 lists the australopith species that will be discussed in this survey.
7.3 Systematic context of australopith diversity

Morphological diversity, taxonomic diversity, and species-lineage diversity do not neatly map onto one another in paleontology. This is largely due to the entrenched use of a phenetic species concept in this science, which tends to endorse the recognition of lineage segments as species. But the delineation of the “actors” in evolutionary processes over geological spans of time requires that we move beyond the recognition of “paleospecies” as static phenetic constructs, useful only as a formal catalogue of diversity, to rendering them as close to the entities that underpin the notion of the genetical species as lineages evincing signs of a unique network of gene exchange (see Kimbel and Rak 1993 for discussion and references). Substantial change can in principle accrue in such lineages under a model of anagenesis—amounts that would prompt many paleontologists to recognize distinct paleospecies at widely separated temporal cross sections of a lineage—but, in the absence of cladogenesis, no actual increase in lineage diversity would have resulted from this process. Therefore, to understand the diversity of australopiths it is one thing to identify and name morphologically distinct taxa and quite another to delineate species–lineages. Although a morphologically diagnosed species and a species–lineage may exactly coincide, where fossil samples with different morphologies and temporal distributions are involved, the identification of species–lineages entails a complex analytical process involving the comparison of phylogenetic hypotheses based on polarized morphological character states (cladistics) with data on the temporal distribution of form (stratophenetics). Where these approaches yield the same character transformations, with minimal stratigraphic

<table>
<thead>
<tr>
<th>Taxon</th>
<th>Geographical distribution</th>
<th>Known age range</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Australopithecus africanus</em></td>
<td>Dart 1925 South Africa</td>
<td>ca. 3.0–2.5 Ma</td>
</tr>
<tr>
<td><em>Australopithecus robustus</em></td>
<td>(Broom 1938) South Africa</td>
<td>ca. 1.5–2.0 Ma</td>
</tr>
<tr>
<td><em>Australopithecus boisei</em></td>
<td>(Leakey 1959) Tanzania, Kenya, Ethiopia, Malawi</td>
<td>≥2.3–1.4 Ma</td>
</tr>
<tr>
<td><em>Australopithecus afarensis</em></td>
<td>Johanson et al. 1978 Ethiopia, Tanzania, Kenya</td>
<td>≥3.7–3.0 Ma</td>
</tr>
<tr>
<td><em>Australopithecus aethiopicus</em></td>
<td>(Arambourg and Coppens 1968) Ethiopia, Kenya, Tanzania(?)</td>
<td>ca. 2.7–2.3 Ma</td>
</tr>
<tr>
<td><em>Australopithecus anamensis</em></td>
<td>Leakey et al. 1995 Kenya</td>
<td>4.2–4.0 Ma</td>
</tr>
<tr>
<td><em>Australopithecus garhi</em></td>
<td>Asfaw et al. 1999 Ethiopia</td>
<td>2.5 Ma</td>
</tr>
<tr>
<td><em>Kenyanthropus platyops</em></td>
<td>M.G. Leakey et al. 2001 Kenya</td>
<td>3.5 Ma</td>
</tr>
</tbody>
</table>
gaps and no autapomorphies in temporally intermediate samples, the simplest explanation of the observed data may be the existence of a temporal sequence of samples representing ancestral and descendant populations within a single lineage (Smith 1994) even if two or more distinct nominal species are thereby joined in one evolutionary species (Simpson 1951; Wiley 1978; Krishtalka 1993). There is always the possibility that the evidence for an anagenetic lineage may mask a poorly resolved record of cladogenesis. Accordingly, a fairly densely sampled fossil record over a well-calibrated stratigraphic record and a reasonably complete sampling of taxa from the relevant time period are necessary for this kind of analysis. These requirements are rarely met in the hominin fossil record (but see Wood et al. 1994 and Kimbel et al. 2006 for exceptions).

In the following survey, I examine australopith diversity as a record of nominal taxa as well as attempting to characterize it in terms of species lineages. Because I am concerned with taxonomic diversity, my review focuses on craniodental morphology.

### 7.4 Species diversity of australopiths

#### 7.4.1 Australopithecus anamensis M. Leakey et al. 1995

*Known temporal distribution:* 4.17 to ca. 3.95 Ma, based on radioisotopic age determinations on tephras (Leakey et al. 1995, 1998)

*Known geographical distribution:* Two sites, Kanapoi, southwest of Lake Turkana, and Allia Bay, east of Lake Turkana, Kenya (Koobi Fora Formation)

*Holotype:* KNM-KP 29281, adult mandible with complete dentition and associated temporal–bone fragment from Kanapoi

*Hypodigm:* Forty-seven catalogued fossils have come from the Kanapoi site (including the distal humerus reported by Patterson and Howells 1967); 31 specimens have come from Allia Bay (Ward et al. 2001). The combined sample comprises mostly dentognathic specimens, including, from Kanapoi, three adult mandibles, an adult maxilla, and deciduous and permanent teeth; from Allia Bay, an adult mandibular fragment, two fragmentary adult maxillae, and deciduous and permanent teeth. Postcranial material includes a distal humerus, a capitate, a proximal hand phalanx, and associated proximal and distal portions of a tibia from Kanapoi. Fragments constituting most of an adult radius from Sibilot Hill (Heinrich et al. 1993), ca. 20 km from the main fossil-bearing locality at Allia Bay, have been attributed to the species based on its inferred geological age (Ward et al. 2001).

*Diagnostic morphology:* Craniodentally, *A. anamensis* differs from extant African apes and Mio-Pliocene hominins (such as *Ardipithecus* and *Sahelanthropus*)
chiefly in its expanded postcanine dental battery, including occlusally more complex premolars and thicker cheek-tooth enamel, characteristics that are shared with later australopiths and early Homo. In morphology and functional wear, *A. anamensis* canines are less apelike than those of earliest known hominins (i.e., *Ardipithecus kadabba*; Haile-Selassie et al. 2004).

Differences from *A. afarensis* and later australopiths pervade the dentition, mostly in the anterior arcade, as well as in the symphyseal region of the mandible and the nasal region of the maxilla. The lower lateral incisors are mesiodistally expanded, as are the lower third premolars and maxillary canines. The P3 is more asymmetric in occlusal form and is uniformly single-cusped, with an incipient metaconid expressed as a tiny pyramidal expansion of the transverse crest rather than a fully developed cusp, and it usually has a relatively large, mesiolingually “open” anterior fovea. Maxillary canine crowns are symmetric in lateral view, with basally positioned mesial and distal shoulders. The buccolingually compressed and occlusally simple deciduous first molar crown differs from the more fully molarized dm1 of subsequent australopiths and recalls the apelike form of this tooth in *Ardipithecus*.

Mandibles (KNM-KP 29281, KNM-KP 29287, KNM-KP 31713) feature an externally convex anterior corpus and a strongly inclined, retreating symphyseal cross section with a “cut away” basal segment, which is manifested in a marked inferomedial inflection of the lower lateral corpus in the canine/premolar region. The lower canine crown is set lateral to the long axis of the postcanine toothrow, giving the dental arch a long, rectangular shape that is reflected in the nearly parallel-sided maxillary dental arch.1 In the maxilla (KNM-KP 29283), large canine roots shape the morphology around the nasal aperture. The lateral margins of the aperture are rounded, not sharp as in *A. afarensis*, and the inferior margin is indistinct, with the subnasal surface arching smoothly into the nasal cavity.

A single fragmentary temporal bone (KNM-KP 29281) forms part of the holotype. From a small evidently female individual, it has a flat mandibular fossa with an indistinct articular eminence, and a horizontally disposed, shallow, and sagittally convex tympanic element that extends laterally to a small-diameter external auditory meatus. Although each of these features can be found individually among *A. afarensis* temporal bones from Hadar, their strong apelike expression in combination is not encountered in the Hadar sample.

---

1 However, based on what little is actually preserved of the midline cross section, the reconstructed dental arch in the maxilla KNM-KP 28283 is too narrow and posteriorly convergent as depicted in published photographs (e.g., Leakey et al. 1995: Figure 1b; partly corrected in Ward et al. 2001: Figure 3).
The tibia and hand phalanx from Kanapoi, and the radius from East Turkana (granting its assignment to *A. anamensis*), are very similar to other australopith homologues, and the tibia constitutes the earliest known skeletal evidence for australopith bipedality. The capitate (KNM-KP 31724) is distinctive, however, in the relative orientation of the distal facets for the second and third metacarpals. As in great apes, but unlike the condition in later australopiths and humans, the facet for MC II is laterally directed and set at approximately 90° to the facet for MC III, implying little or no rotational capability at the carpal/MC II joint (Leakey et al. 1998).

**Discussion:** The *A. anamensis* hypodigm presents a distinctive and significantly more apelike anatomical package than those of subsequent australopith species. However, when the sample is separated into subsamples by site, a more complex picture emerges. Although the younger sample from Allia Bay is much smaller and less representative anatomically than the one from Kanapoi, in several of its morphological details it stands out relative to conditions in the latter sample (Kimbel et al. 2006). The Allia Bay P₃ (in mandible KNM-ER 20432) has an expanded posterior fovea, and its anterior fovea is partly sealed by an elevated mesial marginal ridge. The mandible corpus fragment in which this tooth is preserved does not appear to possess the inferomedial inflection of the lateral corpus under the premolars, implying a less retreating symphyseal cross section. The Allia Bay lower canines (*n* = 2) show less development of the distal cingulum than in the Kanapoi sample (*n* = 4). These distinctions of the Allia Bay sample are the most common states in *A. afarensis* and bridge *A. anamensis* to that taxon. Moreover, as suggested initially by Leakey et al. (1995), the differences between *A. anamensis* and *A. afarensis* are not as pronounced when the geologically older Laetoli sample of *A. afarensis* is considered separately from the younger sample from Hadar (see below, under *A. afarensis*). In fact, the hypothesis of anagenetic change efficiently accounts for the observed morphological transformations across the four temporally ordered site–samples, implying that *A. anamensis* and *A. afarensis* constitute a single evolutionary species (Kimbel et al. 2006).


### 7.4.2 *Australopithecus afarensis* Johanson et al. 1978

**Known temporal distribution:** 3.7–3.0 Ma based on radioisotopic age determinations on tephras
**Known geographical distribution:** Laetoli, Tanzania (upper Laetolil Beds); Hadar, Ethiopia (Sidi Hakoma, Denen Dora and Kada Hadar Members, Hadar Formation); Dikika, Ethiopia (Basal Member, Hadar Formation), Maka, Middle Awash, Ethiopia (“Maka Sands,” Matabaietu Formation), East Turkana, Kenya (Tulu Bor Member, Koobi Fora Formation), (?)Koro Toro, Bahr-el-Ghazal, Chad

**Holotype:** LH-4, mandible with dentition, from upper Laetolil Beds, Laetoli, Tanzania

**Hypodigm:** Approximately 90% of the hypodigm of *A. afarensis* comes from Hadar Formation sediments exposed at the Hadar site in Ethiopia (*n* = 367 specimens). The Hadar sample includes two partial skeletons with craniodental associations (A.L. 288-1, A.L. 438-1), two nearly complete (A.L. 444-2, A.L. 822-1) and one partial (A.L. 417-1) skull, 57 adult or subadult mandibular portions, 13 adult or subadult maxillae, 12 calvarial specimens (including a partial juvenile cranium, A.L. 333-105), and a wealth of upper limb, lower limb, and axial material, the majority of which comes from the A.L. 333 locality (see descriptive papers in the March 1982 *American Journal of Physical Anthropology* and Latimer et al. 1987; Latimer and Lovejoy 1989,1990a, b; Kimbel et al. 1994, 2004; Drapeau et al. 2005). The Laetoli site–sample consists of the holotype mandible (LH-4), several adult or subadult jaws (Garusi I, LH-2, LH-5), a fragmentary partial skeleton of a juvenile (LH-21), and an assortment of teeth (White 1977, 1980, 1981). Trace fossils in the form of the Laetoli bipedal hominin footprint trails, presumed to have been made by *A. afarensis*, are part of the species’ hypodigm. From the Maka site in Ethiopia have come a proximal femur, a humerus, and several jaws, including a nearly complete mandible, at 3.4 Ma (White et al. 1993, 2000; Lovejoy et al. 2002). A partial calvaria (KNM-ER 2602; Kimbel 1988) is known from Koobi Fora. Although the 3.0- to 3.5-yr-old mandible fragment (KT-12/H1) from Koro Toro, Chad, has been made the holotype of *A. bahrelghazali* by Brunet et al. (1996), its anatomy does not appear to distinguish it from the range of variation encompassed by the Hadar/Maka mandibular series, and so it is here tentatively interpreted to extend the geographical distribution of *A. afarensis* into north-central Africa. The Middle Awash frontal fragment from Belohdelie may also belong to *A. afarensis*, which would extend the time range of the species back to ca. 3.9 Ma (Asfaw 1987; Kimbel et al. 2004).

**Diagnostic morphology (Figure 7.1):** *A. afarensis* can be distinguished from *A. anamensis* by a more molarized, symmetrical P$_3$ crown with more frequent development of the second cusp (metaconid) and transverse orientation in the toothrow; an asymmetric upper canine with more apically positioned mesial crown shoulder; a molarized dm$_1$ with buccolingually expanded talonid; a straighter, commonly more upright anterior corpus profile, with a “filled out” basal segment and little to no inferomedial inflection of the lower corpus beneath
the canine–premolars; lower canines set medial to postcanine row axis; a wider palate at equivalent palate lengths; a nasal aperture defined by sharp lateral margins and distinct inferior margin, with the canine jugum a distinct entity in the circum-nasal topography; and a larger external auditory meatus. In all of these features, *A. afarensis* exhibits the derived condition for hominins.

Compared to all subsequent australopiths, a distinctive, predominantly plesiomorphic, anatomy pervades the *A. afarensis* skull and dentition. The calvaria testifies to an extensive posterior *m. temporalis* origin, with posteriorly extended sagittal crests and compound temporal/nuchal crests in both large (putatively male) and small (putatively female) specimens. The nuchal plane of the occipital bone is transversely convex and set at a steep angle to the Frankfurt plane, especially in less heavily crested small and subadult individuals, and it transitions to a relatively short occipital plane. The cranial base features a shallow mandibular fossa bounded anteriorly by a weak to moderately developed

---

Figure 7.1
Three-quarters view of the reconstructed skull of *A. afarensis* specimen A.L. 444-2 from Hadar, Ethiopia. (Photo by W. Kimbel and Y. Rak)
articul ar eminence and posteriorly by an inflated postglenoid process that sits anterior to the tympanic element. The tympanic is a horizontally inclined tube, rather than a vertically oriented plate, and it usually lacks a distinct crista petrosa. Venous blood outflow was predominantly through the occipital–marginal rather than through the transverse–sigmoid sinus system. The frontal bone features a low, flat to mildly convex squama lacking a frontal trigone, coronally oriented, laterally thickened supraorbital bars, and broad postorbital breadth relative to other facial breadth dimensions. The midfacial axis (nasion-nasospinale) is upright, in contrast to the strongly projecting, convex subnasal plane (nasospinale-prosthion), which protrudes anteriorly beyond the bicanine line (also seen in *A. garhi*). Interorbital and nasal aperture breadths are narrow, contrasting with the broad, flat, and massive zygomatic region. The maxilla’s zygomatic process root is located above $M^1$ or $P^4–M^1$, and its inferior margin is moderately to strongly arched. Within the nasal cavity, an elevated platform separates the inferior nasal margin from the anterior vomeral insertion. The palate is moderately deep in narrow jaws but is shallower in wider ones, and the upper dental arch is subparallel with slight convergence to moderate divergence posteriorly. The mandible has a deep but relatively narrow corpus at the molars, and a moderately tall ramus originating high on the corpus and separated from it by a narrow extramolar sulcus. In individuals with an associated mandible and maxilla (i.e., A.L. 417-1, A.L. 444-2, A.L. 822-1), the mandibular corpus constitutes more than two-thirds of the orbitoalveolar height in the coronal plane of the orbits. Although the symphyseal axis is less inclined than in *A. anamensis*, it is, on average, less vertical than in subsequent australopiths. The lower anterior dental arch is pinched in small mandibles but widens into a U-shape in larger specimens, perhaps under the influence of moderate canine dimorphism.

Most of the distinctive dental features of *A. afarensis* are focused in the canines and premolars, which in this species are captured in evolutionary transition. Molarization of the $P_3$ is less advanced than in subsequent australopiths: the metaconid is variably expressed, with some individuals primitively lacking a distinct second cusp (e.g., A.L. 128-23, 277-1, 288-1, 417-1), though these are not necessarily those with a more “apelike” skewed occlusal outline and oblique orientation in the toothrow. Absolute canine size overlaps that of *A. africanus*, though relative to postcanine size the *A. afarensis* canines are larger. Some canines reveal ancestral traces of shearing wear (e.g., on the elongated distal crests of the lowers), which is lost almost entirely in subsequent australopith species, but this is significantly less developed than in *A. anamensis* and especially *Ardipithecus*, and occlusal wear in *A. afarensis* canines and premolars is predominantly apical. Nevertheless, in some mandibles canine and mesial $P_3$ crowns stand tall even in
the face of extreme occlusal molar wear, a remnant of the ancestral occlusal wear pattern.

Discussion: Well-documented variation in the large A. afarensis sample indicates high levels of cranial polymorphism. Some of this variation is due to sexual dimorphism in size and shape, while some of it is due to anagenetic trends in craniodental morphology during the younger half (3.5–3.0 Ma) of the A. anamensis–A. afarensis species–lineage (Lockwood et al. 2000). The Laetoli sample (ca. 3.7–3.5 Ma) figures prominently in this discussion. Although some dental metric differences between the Laetoli and Hadar samples have been cited in the past (White 1985), it now appears that the metric and morphological differences between the hominin populations represented at these sites are phylogenetically significant. Despite limited samples, the convex, retreating form of the adult anterior mandibular corpus (LH-4), mirrored in that of a juvenile (LH-2), the influence of the canine root on the curved lateral margin, and the indistinct inferior margin, of the nasal aperture (Garusi I), and mesiodistally expanded lower canine and P3 crowns recall conditions diagnostic of A. anamensis and occur in more derived states in the Hadar sample (Kimbel et al. 2006).

Later in time, in the Kada Hadar Member at Hadar (ca. 3.0–3.1 Ma), there is evidence of an anagenetic increase in hominin skull size (and perhaps body size), which drove the A. afarensis mandibular corpus and facial skeleton to sizes rarely encountered in older strata (Lockwood et al. 2000). This trend was not accompanied by especially large postcanine teeth, more robust mandibular corpora (i.e., thicker in relation to depth), or by anteriorly shifted masticatory muscle blocks that are usually thought to signal diet-related morphological specializations in subsequent species of the australopith group. The causal processes underlying the observed increase in size are presently unclear, but it does roughly correspond to a change in the Hadar mammalian faunal community to more arid-adapted taxa (e.g., among the bovids).


7.4.3 Kenyanthropus platyops M.G. Leakey et al. 2001

*Known temporal distribution:* ca. 3.3–3.5 Ma based on radioisotopic dating of tephra

*Known geographic distribution:* Lomekwi (Nachukui Formation), west of Lake Turkana, Kenya
**Holotype:** KNM-WT 40000, a crushed and distorted cranium with partial dentition

**Hypodigm:** In addition to the holotype, a partial maxilla with dentition (KNM-WT 38350). (Other hominin fossils, including a partial temporal bone, a mandible fragment [formerly attributed to *A. afarensis* by Brown et al. 1993], and some isolated teeth, were withheld from attribution to this taxon by Leakey et al. [2001].)

**Diagnostic morphology:** While enumeration of the diagnostic features of *K. platyops* is hampered by poor preservation of the holotype, taphonomy does not appear to account for all of the distinctive morphology of the type cranium relative to that of known specimens of *A. afarensis*, with which it was contemporary. These differences reside mainly in the lower part of the face (Leakey et al. 2001) and include an anteriorly positioned root of the maxillary zygomatic process (above P3–P4, and seen in both the type specimen and KNM-WT 38350) and a transversely and sagittally flat subnasal plane with minimal projection beyond the canines. In addition, the dominant venous outflow track was via the transverse–sigmoid system, in contrast to the occipital–marginal system, which is very common in *A. afarensis* (though the Laetoli juvenile LH-21 also evinces the transverse–sigmoid drainage route). The size of the external auditory meatus is smaller than that of *A. afarensis*, despite the latter’s extensive variation in this feature (Kimbel et al. 2004) and approaches the tiny EAM of *Ardipithecus ramidus* and *Au. anamensis*. The second molar in the type cranium is much smaller than that of other australopith M2’s, while the M1 of KNM-WT 38350 is also very small.

**Discussion:** The specimens attributed to *K. platyops* currently constitute the best evidence for hominin lineage diversity prior to 3.0 Ma. On the basis of published information (Leakey et al. 2001), the material attributed to *K. platyops* shares only primitive characteristics with *A. afarensis* (e.g., tubular tympanic element lacking a petrous crest, posteromedially angled temporal lines and an emphasis on posterior cranial crests, low and curved zygomaticoalveolar crest; see Leakey et al. 2001), but its facial configuration appears more derived than what is observed in the large cranial sample of that species. None of these derived characteristics is observed in the sample of one dozen adult maxillae in the Hadar sample of *A. afarensis*, or in the Garusi maxilla from Laetoli, which is demonstrably more primitive in the circumnasal region than the Hadar specimens (as discussed above and in greater detail in Kimbel et al. 2006) yet is the more precise chronological match for the *Kenyanthropus* holotype. (KNM-WT 38350 is approximately contemporary with Hadar specimens A.L. 417-1, A.L. 200-1 and others from the middle Sidi Hakoma member.) Therefore, *K. platyops*
sits in phylogenetic isolation. There are no shared derived characters linking it to the \( A. \ anamensis \)-\( A. \ afarensis \) species–lineage, and although favorable comparisons have been made directly with the lower facial morphology of \( Homo \ rudolfensis \) (i.e., KNM-ER 1470) at ca. 2.0 Ma (Leakey et al. 2001), it is unrealistic to link these two specimens via a meaningful phylogenetic hypothesis given 1.5 Myr of no intervening data.

**Key References**: M.G. Leaky et al. 2001.

### 7.4.4 *Australopithecus africanus* Dart 1925

**Known temporal distribution**: ca. 3.0–2.5 Ma, based on biochronological correlations with radioisotopically calibrated sequences in eastern Africa

**Known geographical distribution**: Taung, Sterkfontein, and Makapansgat, South Africa

**Holotype**: Juvenile skull, dentition, and endocast from Taung

**Hypodigm**: In addition to the holotype, crania, jaws, teeth, and postcrania from Sterkfontein and Makapansgat. The Sterkfontein sample is by far the more extensive, consisting of some dozen nearly complete or partial crania (e.g., TM 1511, Sts 5, 17, 71, Stw 13, 505), adult and juvenile mandibles plus hundreds of teeth, in jaws or isolated. Numerous postcranial remains are known, including at least three partial skeletons (Sts 14, Stw 431, Stw 573: Toussaint et al. 2003), but none is definitively associated with taxonomically diagnostic craniodental remains. The Makapansgat sample comprises roughly 40 specimens, including several jaws (MLD 2, 6/23, 9, 18, 40), two partial adult calvariae (MLD 1, MLD 37/38), and some fragmentary cranial and postcranial elements.

**Diagnostic morphology (Figure 7.2)**: Compared to \( A. \ afarensis \), \( A. \ africanus \) has a higher, shorter braincase with rare sagittal cresting and no compound temporo-nuchal cresting. The supraorbital torus thins laterally from top to bottom and is occasionally divided into distinct supraorbital and superciliary components (Lockwood and Tobias 1999). The cranial base is narrower relative to its length and the foramen magnum is located more posteriorly. The occipital plane of the occipital bone is higher and the nuchal plane is flatter and more horizontally inclined. The mandibular fossa is deeper on average, with a stronger articular eminence. The tympanic element is more vertically oriented, usually bears a distinct crista petrosa, and tapers medially to a distinctive eustachian process. Venous drainage from the endocranium is predominantly through the transverse–sigmoid system. Prominent anterior pillars border the nasal aperture, even in young individuals, and the subnasal plate is flat to slightly convex sagittally and much less projecting relative to the bicanine axis. Within the nasal cavity, the step
down to the nasal floor and anterior vomeral insertion occurs immediately posterior to nasospinale, usually without an intervening platform. The zygomatic bone features a variably prominent boss at the transition to the temporal process and a strong sagittal inflexion across the frontal process/facial surface transition, which combine to create a central facial hollow in some individuals (e.g., MLD 6/23, TM 1511, Sts. 71, Stw. 505). Zygomatic process roots typically originate more anteriorly (above P⁴/M¹ to P⁴/P³) and have a straight, superolaterally diverging inferior margin. The palate is, on average, deeper, with an inferiorly flexed premaxillary segment and posteriorly divergent toothrows. The mandible corpus is more robust (breadth as a percentage of depth at M₁) and has a more inflated lateral surface beneath the premolars and a straighter, more vertical symphyseal profile. Dentally, compared with A. afarensis, A. africanus has absolutely larger (especially broader) postcanine teeth with centrally crowded molar cusp apices in some individuals. The P₃ is uniformly bicuspid, canine wear is exclusively apical, and the anterior–posterior adult occlusal wear gradient is weaker. C/P₃ and I²/C diastemata are less frequent.

In its moderate (though impressively variable) upper midfacial prognathism (nasion-nasospinale), occasionally patent premaxillary suture in adult faces, and less topographically complex glenoid region of the temporal bone, A. africanus remains plesiomorphic in relation to A. robustus and A. bosei (but not
A. aethiopicus). The mandibular corpus is less robust compared to states observed in A. aethiopicus, A. robustus, and A. boisei, and dental size and the anterior–posterior tooth row proportions remain relatively conservative (especially those involving the premolars, which are less expanded in relation to the molars). Facial breadths are narrower as a percentage of calvarial breadths, giving the impression that the cranium is less constricted postorbitally, though the minimum frontal breadth dimension is of similar absolute magnitude. The vertex of the braincase (in specimen Sts. 5 at least) is higher in relation to the orbital roof when compared to crania of A. aethiopicus, A. robustus, and A. boisei (as well as A. afarensis and the great apes), but when size-standardized by calvarial length, vertex height is greater only than that of the “robust” species.

Discussion: The large A. africanus sample from the Sterkfontein Member 4 (“Type Site”) fossil assemblage is notable for its unusually extensive range of craniodental variation (e.g., basal aspect of the temporal bone, facial topography and subnasal prognathism, postcanine tooth size), which has led to proposals that more than one hominin species may be contained within it (Clarke 1988, 1994; Kimbel and White 1988; Kimbel and Rak 1993; Moggi-Cecchi et al. 1998; Lockwood and Tobias 2002). However, because dividing the sample to create overlapping cranial and dental morphs has proven difficult (while different authors tend to disagree on such divisions in the first place), these proposals have not attracted much support. Still, appeals to normal sources of variation within a single geographically delimited biological species (chiefly sexual dimorphism) do not account for the variation, as much of this does not follow patterns of dimorphism observed within extant catarrhine species (Kimbel and White 1988; Kimbel and Rak 1993). The lack of chronological control within the Member 4 faunal assemblage leaves open the possibility that the hominin sample is time-transgressive. Kimbel (1986) offered as an explanation for the Sterkfontein Member 4 variation, the temporal mixing of individual organisms’ remains from morphologically distinct populations of A. africanus that had moved in and out of the Sterkfontein valley over a long period of time, and though Kuman and Clarke’s (2000) revised stratigraphy implies a long and complex depositional history for Member 4, this idea is difficult if not impossible to test, and does not in any event preclude the inclusion of fossils from more than one hominin species in the Member 4 sample. The basic systematics of the type species of Australopithecus remains a vexing issue, and the likelihood of being able to discern lineages is remote.

More recently, excavations in previously unexploited depositories in the Sterkfontein cave system have led to discoveries of hominin fossils said to be significantly older (ca. 4 Ma) than A. africanus of Member 4 (Clarke 1998, 2002; Partridge et al. 2003). These remains include a partial skeleton and skull (Stw 573)
still being excavated in situ in the Silberberg Grotto, and a small collection of specimens, including fragments of an adult cranium (Stw 578), from Jacovec Cavern. Published reports suggest general australopith affinities for these specimens. Their exact antiquity is uncertain.

**Key References:** Broom and Schepers, 1946; Broom et al., 1950; Robinson, 1956; White et al., 1981; Rak, 1983; Clarke, 1994; Lockwood and Tobias, 1999, 2002.

### 7.4.5 *Australopithecus aethiopicus* (Arambourg and Coppens 1968)

**Known temporal distribution:** ca. 2.3–2.7 Ma

**Known geographic distribution:** Omo River basin, Ethiopia (Shungura Formation, Members C–F); West Turkana, Kenya (Nachukui Formation); Laetoli, Tanzania (Ndolanya Beds)

**Holotype:** Omo 18-1967-18, edentulous mandibular corpus (Member C, Shungura Formation, Ethiopia, ca. 2.7 Ma)

**Hypodigm:** The mostly complete and largely edentulous cranium KNM-WT 17000 from the Nachukui Formation, West Turkana, Kenya (2.5 Ma; Walker et al. 1986; Leakey and Walker 1988) is by far the most complete and well-preserved evidence for this relatively poorly known australopith species. Suwa et al. (1996) assigned to this species some 20 lower postcanine teeth (several in mandibles) from Omo Shungura Formation Members C–F, aggregate spanning ca. 2.7–2.3 Ma. An edentulous maxilla (EP 1500/01) from the Ndolanya Beds at Laetoli has also been assigned to *A. aethiopicus* (Harrison 2002).

**Diagnostic morphology:** Based on the single known cranium (KNM-WT 17000), *A. aethiopicus* can be distinguished from *A. afarensis* by its extreme midfacial prognathism (nasion-nasospinale); flat subnasal plane; smooth transition from subnasal surface to nasal cavity floor, with indistinct inferior margin of nasal aperture; anterior vomeral insertion and anterior nasal spine merged within nasal cavity; vertically thick palate; anteriorly positioned zygomatic process roots (over P³/P⁴); bulbous, forwardly sloping zygomatic facial surface, which leaves the nasal region in a central facial hollow; low calvarial height; frontal squama with frontal trigone delimited by strongly convergent temporal lines; vertically inclined tympanic element with distinct petrous crest; coronally aligned petrous element; massive postcanine dentition (inferred from roots and fragmentary P⁴ crown; Suwa 1989). Although the KNM-WT 17000 cranium features a very high sagittal crest, its diagnostic value is doubtful, as it was produced by the
juxtaposition of enlarged temporalis muscles to a very small braincase (410 cm³ cranial capacity) that may not have been typical for the species.

The majority of the characters discriminating the *A. aethiopicus* cranium from that of *A. afarensis* link the former species to classical “robust” australopith masticatory configurations. In addition, the dental remains attributed to the former species testify to postcanine size expansion and molarization seen in otherwise only in *A. robustus* or *A. boisei*, though with less premolar crown specialization than in the latter taxon (Suwa 1990; Suwa et al. 1996). However, *A. aethiopicus* retains a number of plesiomorphic characters, such as strong midfacial prognathism, posteriorly accentuated sagittal crest and extensive compound temporonuchal crest, flat mandibular fossa with low, indistinct articular eminence, and (as inferred from roots and alveolar dimensions) relatively large incisors and/or canines. While these characters set *A. aethiopicus* apart from both *A. robustus* and *A. boisei*, they also effectively discriminate it from *A. africanus*, which in several of these respects exhibits derived morphology (cranial vault shape, crest configuration, and mandibular fossa topography).

**Discussion:** As noted in the historical survey earlier, the discovery of the KNM-WT 17000 cranium provoked discussion about the role of homoplasy in early hominin phylogeny (see contributions in Grine 1988). The fact that aspects of this specimen’s cranial morphology are strikingly more plesiomorphic than homologous states in *A. africanus*, combined with its obvious “robust” australopith masticatory signal, implies a high degree of convergent evolution in the australopith skull, the nature of which depends to a large extent on the phyletic position accorded *A. africanus*.

Does the temporal juxtaposition of the morphologically intermediate *A. aethiopicus* between *A. afarensis* and *A. boisei* support the identification of an evolving australopith lineage (evolutionary species) in the Middle to Late Pliocene of eastern Africa? Walker et al. (1986; Leakey and Walker 1988) thought so and interpreted KNM-WT 17000 as an early, primitive *A. boisei* specimen. Suwa’s (1990; Suwa et al. 1996) detailed analysis of australopith postcanine dental evolution in the Shungura Formation found that in most respects the premolars and molars from Members C–F he attributed to *A. aethiopicus* more closely approximate the generalized condition of *A. robustus* than the highly derived condition of *A. boisei* (e.g., Olduvai Gorge, Koobi Fora, Peninj, etc.), but that the “*A. boisei* morphology emerge[d] in a mosaic fashion across Member G times [i.e., <2.3 Ma]” (Suwa et al. 1996 p 274), implying in situ evolution of this morphotype. Even so, phylogenetic analyses of cranial morphology have usually identified a monophyletic *A. robustus + A. boisei* clade (Strait et al. 1997; Kimbel et al. 2004) whose last common ancestor would be expected to be cranially considerably more derived than the KNM-
WT 17000 specimen of *A. aethiopicus*. Thus, given the still spotty fossil evidence in the 2.8–2.3 Ma time interval, the existence of a monophyletic east African “robust” lineage stemming anagenetically from *A. afarensis* remains weakly supported.

**Key References**: Walker et al., 1986; Leaky and Walker, 1988; Grine, 1988; Kimbel et al., 1998, 2004; Suwa et al., 1996.

### 7.4.6 *Australopithecus garhi* Asfaw et al. 1999

**Known temporal distribution**: ca. 2.5 Ma based on radioisotopic dating of tephra

**Known geographical distribution**: Bouri, Middle Awash, Ethiopia (Hata Member, Bouri Formation)

**Holotype**: BOU-VP 12/130 partial cranium, consisting of frontal, parietal, and maxilla with dentition.

**Hypodigm**: As for holotype.

**Diagnostic morphology**: The single specimen of *A. garhi* combines a largely plesiomorphic face and palate with a derived calvarial morphology and a highly unusual hominin dental configuration (Asfaw et al. 1999). It shares with *A. afarensis* a remarkable number of maxillary features including a strongly prognathic, convex subnasal surface; sharp lateral margins of the nasal aperture due to lateral placement of the canine root jugum and lack of anterior pillar; horizontal separation of the vomeral and septal insertions within the nasal cavity; low, arched root of the maxillary zygomatic process, and shallow palate. Although the parietal bones of this apparently male individual bear a fused sagittal crest, it does not exhibit the strong posterior emphasis seen in larger *A. afarensis* crania (e.g., A.L. 333-45, A.L. 444-2). The anterior part of the frontal squama features a depression bounded by the convergent temporal lines, said to be frontal trigone (Asfaw et al. 1999), which is otherwise observed only in *A. aethiopicus*, *A. robustus*, and *A. boisei*.

Dentally, the *A. garhi* cranium combines huge postcanine teeth with large incisors and especially canines. Although postcanine size falls above even the *A. boisei* range, neither enamel thickness nor occlusal form aligns these teeth with “robust” australopiths. The ratio of the calculated crown area of the *A. garhi* P3 to that of M2 (70%, according to data in Asfaw et al. 1999) is greater than that of either *A. afarensis* (mean = 55%, range = 51–61%, n = 5 individuals with both teeth) or *A. africanus* (mean 54%, range = 50–61%, n = 8 individuals with both teeth), and even exceeds this ratio in the measurable sample of *A. robustus* specimens (mean = 59%, range = 52–65%, n = 7), one hallmark of which is relative premolar enlargement. Relative P4 size (P4/M2 ratio = 70%) is essentially
identical to the *A. robustus* mean (69%). Relative anterior tooth size (as measured by the I₁–C alveolar length as a percentage of the square root of various post-canine tooth crown areas) is smaller than in *A. afarensis* (and in early *Homo*), larger than in *A. robustus* or *A. boisei*, and most similar to that of *A. africanus* (Asfaw et al. 1999).

Discussion: The *A. garhi* cranium presents a unique amalgam of plesiomorphic (*A. afarensis*-like) and derived (*A. africanus* or “robust” australopithecine-like) characters. It is similar in this respect to *A. aethiopicus*, although the character combinations are quite distinct in these two contemporary taxa (albeit each represented by only a single cranium). Whereas *A. aethiopicus* facial and premolar crown form are the most derived aspects of this species’ anatomy (notwithstanding its unparalleled degree of midfacial prognathism), in *A. garhi* it is the morphology of the calotte (including the frontal) and premolar–molar proportions that are apomorphic—maxillofacial morphology remains primitive in this taxon. If we grant that *A. africanus* was approximately synchronic with these species, then this early interval of the late Pliocene presents an impressive and, up to this point in geological time unprecedented, array of hominin craniofacial and dental form, much of which appears related to the elaboration of postcanine megadonty subsequent to *A. afarensis*. While the relatively primitive masticatory system of *A. afarensis* makes it a plausible ancestor for any or all of these younger taxa, neither sample size, nor anatomical representation, nor stratigraphic density of the fossil record, make defining lineages across the 3.0- to 2.5-Ma time period a meaningful exercise. Nevertheless, by 2.8 Ma or so, we can perceive the diverse products of what can be hypothesized as a series of speciation events that laid the groundwork for the evolutionary events of the late Pliocene.

Key References: Asfaw et al. 1999.

### 7.4.7 *Australopithecus robustus* (Broom 1938)

**Known temporal distribution:** ca. 1.5–2.0 Ma based on biochronologic correlation with radioisotopically calibrated east African sequences

**Known geographic distribution:** Kromdraai, Swartkrans, Drimolen, South Africa

**Holotype:** TM 1517 cranium, mandible, and purportedly associated talus from Kromdraai

**Hypodigm:** Approximately 250 specimens of jaws, teeth, crania, and postcranial elements from the Swartkrans cave (Brain 1993); a much smaller sample of teeth (approximately 30), in addition to the holotype, from Kromdraai
(Thackeray et al. 2001); and more than 80 dentognathic specimens, including a cranium with associated mandible, from Drimolen (Keyser 2000; Keyser et al. 2000).

Diagnostic morphology (Figure 7.3): Broom’s (Broom and Schepers 1946) inference regarding the specific distinctiveness (from *A. africanus* sensu lato) of

![Figure 7.3](image)

**Figure 7.3**

Anterior view of Swartkrans *A. robustus* cranium SK 48. (Photo by W. Kimbel)

the Kromdraai skull and dentition was based on its disproportionately large postcanine dentition and unusual facial morphology, which features broad, anteriorly placed zygomatic bones and depressed infraorbital surfaces and nasal bridge between them (the so-called “dished face”). The large sample of teeth, jaws, and crania accumulated during postwar excavations at Swartkrans added significantly to the *A. robustus* diagnostic profile, including buccolingually expanded and molarized premolars and lower deciduous first molars; premolar and molar crowns with convergent cusp apices and very thick enamel caps that wear to flat occlusal planes soon after eruption; mandibles with transversely thick and inflated bodies and vertically high rami; deep, posterosuperiorly sloping palatal roofs; a depressed posttoral frontal squama delimited by strongly incurv...
temporal lines (“frontal trigone”); and frequent frontoparietal sagittal crests. The large series of juvenile mandibles (with fully deciduous or mixed dentition) and a fragmentary, edentulous juvenile maxilla (SK 66), all from Swartkrans, demonstrate that many components of this pattern, especially postcanine megadontia and gnathic robusticity were manifested early in craniofacial ontogeny. Robinson (1953b) drew attention to the unusual nasal cavity configuration in \( A. \) robustus, which includes a smooth (“unstepped”) transition from the subnasal surface to the nasal cavity floor, and a retracted position of the anterior nasal spine within the nasal cavity where it merges with the anterior insertion of the vomer (also observed in \( A. \) boisei and \( A. \) aethiopicus).

Rak’s (1983) study of the australopith face identified a constellation of unique circumnasal (nasoalveolar gutter, maxillary fossula—the highly modified canine fossa, anterior pillar—less prominent than what is seen in \( A. \) africanus) and infraorbital traits (zygomaticomaxillary step, maxillary trigon) related to the posterior retraction of the palate and the anterior encroachment of the zygomatic bones (to originate above the premolars) on the midface. These antagonistic “shifts” lend the Swartkrans and Kromdraai crania their humanlike orthognathic appearance, and though the subnasal maxilla itself remained primitively prognathic, the facial aspect of the premaxillary suture fused prior to adulthood, again, as in humans.

Among the humanlike characteristics of the \( A. \) robustus type specimen that most impressed Broom (Broom and Schepers 1946) was the basal aspect of the temporal bone. The variably deep mandibular fossa with discrete articular eminence, small postglenoid process, vertically oriented, platelike tympanic element with prominent petrous crest and frequently well developed vaginal process of the styloid, and the narrow or absent gap between the postglenoid process and tympanic are found with remarkable consistency in the now-numerous cranial bases of \( A. \) robustus (e.g., SK 46, SK 47 [subadult], SK 48, SK 49, SK 52/SKW 18, SK 83, SKW 11, SKW 2581) and all converge on modern human morphology. Though elements of this pattern can be found in \( A. \) africanus, they are not presented in that taxon as a package. The single cranium of \( A. \) aethiopicus (KNM-WT 17000) presents notably less derived morphology in this area, sharing with \( A. \) robustus only the vertical tympanic with a strong petrous crest. The often dramatic lateral extension of the tympanic’s inferior margin beyond the sagittal plane of porion and the coincident expansion of the diameter of the external auditory meatus are not seen in \( A. \) afarensis, \( A. \) africanus, or \( A. \) aethiopicus.

**Discussion:** The craniodental remains of \( A. \) robustus present a coherent and highly distinctive anatomical package across the three South African sites in which they are represented. Although a species-level distinction between the Kromdraai (\( A. \) robustus) and Swartkrans (\( A. \) crassidens) site samples has been
proposed on the basis of metrical and morphological differences in the permanent and deciduous postcanine teeth—according to which the Kromdraai specimens appear less highly differentiated and thus more similar to the teeth of *A. africanus* (Grine 1993)—reported variation within the Drimolen “robust” australopith sample spans the differences between the Kromdraai and Swartkrans dental collections (Keyser et al. 2000), making a site-based taxonomic split difficult to support.

With the proliferation of the east African hominin fossil record in the 1960s–1980s, there developed a tendency to see the South African australopiths as poised in morphological and adaptive transition between the generalized *A. afarensis* and the specialized *A. bosei*. Rak’s (1983) morphcline of australopith craniofacial morphology was explicit in this regard, though when translated into a phylogenetic scenario, it allowed for cladogenesis between an *A. africanus–A. robustus* lineage and *A. bosei* on the basis of the overlap of the two “robust” species’ temporal ranges. The recognition of the biological validity of *A. aethiopicus* rendered the *A. africanus–A. robustus* lineage unlikely due to the extensive character reversal it would entail given the former species’ derived zygomatico-maxillary and postcanine dental morphology. While a polyphyletic origin for the “robust” morphology via separate eastern (*A. aethiopicus–A. bosei*) and southern (*A. africanus–A. robustus*) lineages is plausible, this scheme does not gain support from phylogenetic analyses of craniofacial characters, which strongly back a monophyletic “robust” clade to the exclusion of *A. africanus*. Thus, the potential for a single australopith species lineage in southern Africa is weak (and notwithstanding the poor chronological resolution within these hominin-bearing deposits).


### 7.4.8 *Australopithecus boisei* (Leakey 1959)

*Known temporal distribution*: ca. 2.4–1.4 Ma, based on radioisotopic dating and tephrostratigraphy

*Known geographic distribution*: Olduvai Gorge (Beds I and II), Tanzania; Peninj (Humbu Formation), Tanzania; Omo River basin (Shungura Formation, Member G), Ethiopia; Konso, Ethiopia; Koobi Fora (Koobi Fora Formation); West Turkana (Nachukui Formation), Kenya; Chesowanja, Kenya

*Holotype*: OH 5, cranium with full dentition, from Bed I, Olduvai Gorge

*Hypodigm*: Aside from the holotype, isolated teeth from Beds I and II, Olduvai Gorge (Tobias 1967); several mandibles, a partial cranium, and many
isolated teeth from the Omo River basin (Shungura Formation, Member G); six complete or partial crania, more than 20 mandibles and several maxillae with and without teeth from Koobi Fora (Wood 1991) a partial subadult cranium from West Turkana (Leakey and Walker 1988) two partial crania from Chesowanja (Carney et al. 1971; Gowlett et al. 1981) a mandible with full dentition from Peninj (Leakey and Leakey 1964); a partial cranium with associated mandible, plus isolated teeth and jaw and cranial fragments from Konso, Ethiopia (Suwa et al. 1997).

Diagnostic morphology (Figures 7.4 and 7.5): A. boisei presents one of the most easily recognized morphological patterns in the early hominin fossil record, due chiefly to uniquely derived states for many of its skull and dental characters. Tobias’s (1967 p 232–233) study of the holotype cranium supported the specific distinctiveness (relative to A. robustus) of the Olduvai hominin, citing, among other features, the absolutely larger (especially broader) postcanine teeth, greater

Figure 7.4
Three-quarters view of the reconstructed cranium of A. boisei holotype specimen OH 5. (Photo by D. Johanson)
disproportion between the postcanine and anterior dentitions, stronger supraorbital torus, shorter foramen magnum, absence of “even the slightest trace of a canine fossa,” structure of the zygomatic region, anterior flexion of the palate, and “evidences of marked parietal lobe and cerebellar expansion.” Except for the foramen magnum and inference of brain expansion, these distinctions reside in the dentognathic complex, and it is here that most subsequent writing on the diagnostic morphology of $A. \text{boisei}$ has been focused. In general terms, $A. \text{boisei}$ has often been characterized as “hyper-robust,” essentially an extreme version of $A. \text{robustus}$. Indeed, much of the specific craniofacial morphology of $A. \text{boisei}$ identified by Rak (1983), such as the elevated infraorbital “visor” (and its effects on the zygomatic arch), the (at least occasionally) inferolaterally sloping and
“twisted” supraorbital bars, and the absence of an anterior pillar and maxillary fossula, were interpreted by him as signaling culminations of adaptive trends associated with masticatory specialization (retraction of the palate and advancement of the zygomatic region) that had already begun in *A. africanus*. The massiveness of *A. boisei* mandibular corpora, which (notwithstanding taphonomic effects on preserved dimensions among some of the largest of these; Silverman et al. 2001) surpasses even the most heavily built *A. robustus* jaws so far known (e.g., KNM-ER 3230 or KGA10-525 versus SK 12), but there is a high degree of variation in corpus size (but less so in corpus shape) among the mandibles of both species.

Discoveries made since OH 5 confirm that the postcanine dentition of *A. boisei* is substantially enlarged relative to that of *A. robustus* (White et al. 1981; Grine 1993). The mandibular P4 is particularly diagnostic of *A. boisei*, not only due to its larger size relative to the molars but also to its derived talonid expansion and molarization (Suwa 1988), also a tendency, though not as consistent, of the lower molars (Suwa et al. 1996). Whether measured as lower canine breadth or lower central incisor–canine alveolar length, the *A. boisei* anterior dentition is smaller relative to the cheek teeth than in *A. robustus* (Suwa et al. 1996). (Upper teeth of *A. boisei* remain poorly known relative to the lowers, a phenomenon associated with the differential taphonomic attributes of the mandible versus the maxilla.)

The cranial base of *A. boisei* has received less attention than the dentognathic part of the skull in systematic studies of *A. boisei*, but the morphology here is no less specialized (Kimbel et al. 2004). In lateral aspect, the mastoid process appears as an equilateral triangle, with an extensive, swollen lateral face that is sharply delimited from the posterior face, mostly hidden in this view, and the mastoid tip lies approximately midway between the projected horizontal line spanning asterion to porion. This configuration is unmatched in the large majority of other *Australopithecus* crania, in which the mastoid is dominated by an extensive posterolateral face with a long, straight posterior margin that converges with a shorter anterior margin to position the tip much closer to porion than to asterion. This is the generalized pattern common to the great apes, *A. afarensis* and *A. africanus*. Some *A. robustus* specimens approach the *A. boisei* anatomy (i.e., SKW 11), though most are more generalized, as is that of the *A. aethiopicus* cranium KNM-WT 17000.

The region around the mandibular fossa is also unique in *A. boisei* among australopith species. The articular fossa is very deep, with the highest point on the ceiling reaching above the Frankfurt plane, and the long steep articular eminence terminates at the posterior edge of the temporal foramen. The preglenoid plane is
usually mediolaterally restricted. The articular eminence twists about its long axis such that medially it faces posteriorly, and the pyramidal entoglenoid process likewise points backward to create a flat “medial glenoid plane” at the medial end of the eminence. The platelike tympanic element is often flattened superiorly and inclined posteriorly, giving the external auditory meatus an oval shape with a diagonal long axis. In some specimens (e.g., KNM-ER 406, OH 5), this inclined tympanic forms more of the ceiling of the mandibular fossa than its posterior wall, and the petrous crest blends with the anterior face of the mastoid process, resulting in obliteration of the tympanomastoid groove. In all of these respects, the cranial bases of *A. robustus* and *A. aethiopicus* are more generalized.

**Discussion:** There is statistical backing for a trend (though not necessarily a gradual one) toward larger postcanine tooth size in *A. boisei* (sensu stricto) between 2.0 and 1.4 Ma (Wood et al. 1994; Lockwood et al. 2000). There is no evidence of change in mandibular corpus size over this interval. Prior to 2.0 Ma, the east African “robust” australopith fossil record is not well documented, but evidence for a species lineage that extends back in time from *A. boisei* focuses on specimens that lack the full suite of derived craniodental characters enumerated above. “Robust” australopith dental remains, especially the lower premolars, from Members D–F of the Shungura Formation (ca. 2.5–2.3 Ma) are less derived than those of post-2.0 Ma *A. boisei* (Suwa et al. 1996), as reviewed above (see discussion of *A. aethiopicus*). Cranial specimens of the “robust” morphotype from the pre-2.0-Ma period are rare, but the lower Member G partial cranium Omo 323-1976-896 (ca. 2.1 Ma), attributed by Alemseged et al. (2002) to *A. boisei*, lacks a number of the cranial autapomorphies that characterize the post-2.0-Ma skulls of this species, including mastoid, tympanic, articular eminence, preglenoid, and entoglenoid characteristics (Kimbel et al. 2004; see also Alemseged et al., 2002). In these respects, the Omo cranium resembles the more generalized *A. robustus*, yet it is less similar to *A. aethiopicus*—the Omo specimen has a deeper mandibular fossa and probably, based on the preserved maxillary fragment, less mid-facial prognathism compared to KNM-WT 17000 (2.5 Ma)—which implies an immediate predecessor of *A. boisei* in eastern Africa that was cranially more like southern African *A. robustus* than like eastern African *A. aethiopicus*. Suwa (1988; Suwa et al. 1996) reported a mix of derived and generalized “robust” australopith postcanine crown morphologies in Member G of the Shungura Formation (ca. 2.3–2.0 Ma), which is consistent with the relatively generalized anatomy of the Omo 323 cranium at ca. 2.1 Ma. This evidence has been used to support an anagenetic derivation of *A. boisei* from a chain of progressively less derived “robust” populations leading back to *A. aethiopicus* at ca. 2.5 Ma or older (Kimbel et al. 1988; Leakey and Walker
1988; Suwa et al. 1996; Alemseged et al. 2002). If true, this would argue for the existence of a single-species lineage of over 1-myr duration, which is about the documented span of the hypothesized A. anamensis–A. afarensis species lineage of the earlier Pliocene (Kimbel et al. 2006). However, given the demonstrated morphological links of the eastern African “robust” specimens between 2.0 and 2.3 Ma to A. robustus, and the report of otherwise unique A. robustus characters within a diagnostic A. boisei morphological milieu in the late (1.4 Ma) partial skull from Konso, Ethiopia (Suwa et al. 1997), reconstructing a direct phyletic link between A. aethiopicus and A. boisei would require positing a cladogenetic event giving rise to a southern African “robust” branch within the 2.3- to 2.0-Ma interval. Thus, an unbroken, anagenetically evolving “robust” australopith species lineage in the eastern African middle-late Pliocene is unlikely.


7.5 Summary and conclusions

Given a strict evolutionary species definition, nominal taxonomic diversity and species–lineage diversity do not necessarily map onto one another in the fossil record. Species–lineages entail statements of ancestry and descent that depend on the consistency of phylogenetic and stratophenetic data. The requirements for identifying species lineages in the fossil record are severe and rarely met in the early hominin record, most often owing to small sample size, under-represented character data, nonrepresentation of rare or short-lived taxa, poor chronological resolution, gaps in the time-stratigraphic framework, or some combination of these factors. Because hypotheses concerning the “bushiness” of the hominin phylogenetic tree depend on the identification of lineages, not phenetically based “paleospecies,” confidence with respect to this issue is not justified for the majority of the hominin fossil record.

The fossil record of australopith evolution in eastern Africa offers two cases in which an approach to this question can be attempted. In one, the evidence, though still imperfect, is consistent with the evolution of A. anamensis into A. afarensis via anagenesis (i.e., without an increase in lineage diversity). The other, the evolution of A. boisei, is almost certain to have entailed a speciation event that gave rise to southern African clade (represented by A. robustus) subsequent to the appearance of A. aethiopicus. The late Pliocene time period in which the latter events transpired (ca. 2.8–2.3 Ma) was one of substantial morphological diversity, high nominal taxonomic diversity, and high probability
of synchronicity among known fossil samples. Although it is not possible to connect these australopith taxa (A. aethiopicus, A. africanus, A. garhi) to particular descendants due to the aforementioned defects in the data base—and indeed, it is not unlikely that some or all of these species are terminal taxa—once it is better known, this period will most likely be seen to have witnessed a previously (and subsequently) unmatched degree of lineage proliferation compared to other parts of the human evolutionary record. The challenge to paleoanthropologists is to devote resources to improving this part of the fossil record and then to create testable phylogenetic and adaptive hypotheses to explain it.

References

Broom R, Robinson JT (1952) Swartkrans ape-man, Paranthropus crassidens. Transv Mus Mem 6
Broom R, Schepers GWH (1946) The South African fossil ape-men, the Australopithecinae. Transv Mus Mem 2
Broom R, Robinson JT, Schepers GWH (1950) Sterkfontein Ape-man Plesianthropus. Transv Mus Mem 4
Robinson JT (1956) The dentition of the Australopithecinae. Transv Mus Mem 9
Siess W, Lapetina EG. 1989. Prostacyclin inhibits platelet aggregation induced by phorbol ester or Ca$^{2+}$ ionophore at steps distal to activation of protein kinase C Ca$^{2+}$-dependent protein kinases. Niochem 258: 57–65
8 Defining the Genus *Homo*

*Mark Collard - Bernard Wood*

**Abstract**

The definition of the genus *Homo* is an important but under-researched topic. In this chapter, we show that interpretations of *Homo* have changed greatly over the last 150 years as a result of the incorporation of new fossil species, the discovery of fossil evidence that changed our perceptions of its component species, and reassessments of the functional capabilities of species previously allocated to *Homo*. We also show that these changes have been made in an *ad hoc* fashion. Criteria for recognizing fossil specimens of *Homo* have been outlined on a number of occasions, but these criteria have generally not been explicitly derived from a genus concept. Rather, the course of action followed by most researchers has been to assign new specimens to *Homo* on the basis of a subset of the diagnostic traits that are considered to be key, and to then redefine the other traits of the genus in the light of the morphological and functional attributes of the new specimens. With a view to moving beyond this approach, in the next section of the chapter we outline six competing proposals for how genera should be defined and consider their impact on the species assigned to the genus *Homo*. Subsequently, we consider the pros and cons of the six genus concepts. We argue that three of them are impractical and/or internally inconsistent and the other three are useful. We go on to suggest that, while there is little to choose between the latter three concepts on theoretical grounds, the one put forward by Wood and Collard (1999) has practical advantages. In the last part of the chapter, we update Wood and Collard’s (1999) review of genus *Homo* in the light of research published since their study appeared. We find that, on balance, the available evidence still supports their suggestion that *Homo* should be reconfigured such that it includes *H. ergaster, H. erectus, H. heidelbergensis, H. neanderthalensis,* and *H. sapiens* but excludes *H. habilis* and *H. rudolfensis*. We also find that the proposed inclusion of the collection of Late Pleistocene specimens from the site of Liang Bua, Flores, in the genus *Homo* as a new species, *H. floresiensis*, is not compatible with Wood and Collard’s (1999) definition of the genus *Homo*.
8.1 Introduction

It is obvious—indeed it is so obvious that it bears repeating—that an understanding of the evolution of genus *Homo* depends, to a considerable extent, on the proper definition of the classificatory categories “genus” and “species” and the correct evaluation of the taxa that are assigned to these categories. In recent years, the species category has received considerable attention from paleoanthropologists (e.g. Tattersall 1986; Turner and Chamberlain 1989; Kimbel and Martin 1993) and there has been a taxonomic rationalization of material previously grouped as “early *Homo*” and “archaic *H. sapiens*” into reasonably robust species groups (Tattersall 1986, 1992; Lieberman et al. 1988; Wood 1991; 1992; Wood et al. 1991; Rightmire 1993, 1996, 1998). In contrast, both the definition of the genus category and the demarcation of the genus *Homo* remain contentious. In this chapter, we outline the main events that have taken place in the taxonomic history of the genus *Homo*. We then evaluate several recent proposals to amend the criteria that are used to assign species to *Homo*. Lastly, we update Wood and Collard’s (1999) review of genus *Homo* in the light of research published since their study appeared.

8.2 Changing interpretations of genus *Homo*

The genus *Homo* was established by Carolus Linnaeus in the 10th edition of his *Systema Naturae*, which was published in 1758. As conceived by Linnaeus, the genus incorporated two species. The name *Homo sapiens* was attached to what Linneaus described as the more diurnal of the two species. Within *H. sapiens*, Linnaeus recognized six groups. Four of these are geographical variants drawn from the four continents known to Linnaeus, namely Africa, America, Asia, and Europe. The other two groups, which Linneaus called “wild men” and “monstrous men” respectively, are of historical rather than biological interest. A similar conclusion probably also applies to Linnaeus’ second species of *Homo*, *H. sylvestris*, also called *H. troglodytes* or *H. nocturnes*, which he suggested is a nocturnal cave dwelling form of human from Java. *H. sylvestris* is now thought to be mythical, although the recent discovery of *H. floresiensis* (see below) raises the possibility that it may have had some basis in fact.

The first fossil species was assigned to *Homo* in 1864. In this year, the Irish anatomist William King referred a partial skeleton that had been recovered in 1856 from the Feldhofer cave in the Neander Valley in Germany to *H. neanderthalensis* (King, 1864). King considered naming a new genus for the Feldhofer skeleton but eventually decided that it was sufficiently similar to *H. sapiens* to
warrant its inclusion within *Homo*. In the same year, George Busk reported to the British Association for the Advancement of Science on what we now know to be a Neanderthal cranium from Gibraltar (Busk 1865). Although Busk acknowledged the strength of the resemblance between the Gibraltar cranium and the one from the Neanderthal Cave, he judged the former to belong to *H. sapiens*, albeit a member of the species that was more similar to living Tasmanians and Australians than to contemporary Europeans. The inclusion of the Neanderthal skeleton within *Homo* expanded the ranges of both the cranial and postcranial morphology of the genus. The morphology of the type specimen, together with evidence gleaned from discoveries made prior to 1856 and thereafter in Western Eurasia, show that Neanderthal crania differ from those of *H. sapiens* in several respects. Typically, they have discrete and rounded supraorbital ridges, faces that project anteriorly in the midline, laterally projecting and rounded parietal bones, a rounded, posteriorly projecting, occipital bone, a derived nasal morphology (Schwartz and Tattersall 1996; but see Franciscus 1999), large incisor teeth, and postcanine teeth with large root canals. Their brains were as large, if not larger, in absolute terms than the brains of modern humans. Postcranial peculiarities of the Neanderthals include limb bones with stout shafts and relatively large joint surfaces, especially well-marked areas for the attachment of a muscle that helps to control movement at the shoulder, and an elongated pubic ramus of the pelvis (Pearson 2000).

The morphological variability of genus *Homo* was further extended between 1908 and 1933 by the addition of a group of specimens that was initially referred to as “archaic” *H. sapiens* but is now more often called *H. heidelbergensis* (Tattersall 1986, 1992; Rightmire 1996, 1998). The type specimen of *H. heidelbergensis* is a mandible that was found in 1907 during excavations to extract sand from a quarry at Mauer, near Heidelberg, Germany (Schoetensack 1908). The next evidence within Europe came in 1933 from a gravel pit at Steinheim in Germany, but in the meantime evidence had also been found at the site of Kabwe in what was then Rhodesia (Woodward 1921) and at the Javanese site of Ngandong (Openoorth 1932). The brain cases of *H. heidelbergensis* are often, but not always, smaller than those of modern humans (e.g., Steinheim), but they are always more robustly built, with large rounded ridges above the orbits and a thickened occipital region. The Mauer mandible has no chin and the corpus is substantially larger than those of modern Europeans. Postcranially, the shapes of the limb bones are much like those of *H. sapiens*, except that the shafts of the long bones are generally thicker, with higher robusticity indices. Schoetensack’s (1908) decision to refer the Mauer mandible to *H. heidelbergensis* altered the interpretation of *Homo* in that it added a taxon with a mandible more primitive than those of either *H. sapiens* or *H. neanderthalensis*. The subsequent addition of the Kabwe
and Ngandong specimens to *Homo* meant that the genus now included a species with a more heavily built cranium than either *H. sapiens* or *H. neanderthalensis*.

The range of morphology within *Homo* was widened again in 1940 when Franz Weidenreich formally proposed that two existing hypodigms, *Pithecanthropus erectus* and *Sinanthropus pekinensis*, should be merged into a single species and transferred to *Homo* (Weidenreich 1940). The name for the new species was *Homo erectus* (Weidenreich 1940). Subsequently the hypodigms of *Meganthropus* (Mayr 1944 p 14; Le Gros Clark 1955 pp 86–87), *Atlanthropus* (Le Gros Clark 1964 p 112), and *Telanthropus* (Robinson 1961; Howell 1978 p 198) were also sunk into *H. erectus*. Compared with *H. sapiens*, *H. neanderthalensis*, and *H. heidelbergensis*, fossils attributed to *H. erectus* have a smaller neurocranium, a lower vault, a broader base relative to the vault, and more complex premolar roots. They also have a substantial and essentially continuous torus above the orbits behind which is a sulcus. There is usually a sagittal torus and an angular torus that runs toward the mastoid process. The occipital region is sharply angulated, with a well-marked supratoral sulcus, and the inner and outer tables of the vault are thickened. Despite the relatively large numbers of crania that had been recovered from Java, China, and elsewhere, relatively little was known about the postcranial morphology of what was to become *H. erectus*. Discoveries from East African sites provided crucial evidence in the form of a pelvis and femur from Olduvai Gorge (OH 28), two fragmentary partial skeletons from East Turkana (e.g., KNM-ER 803 and 1808), and the unusually well-preserved skeleton from West Turkana (KNM-WT 15000). The cortical bone of the postcranial skeleton is generally thick. The long bones are robust, and the shafts of the femur and the tibia are flattened from front to back relative to those of other *Homo* species; these conditions are referred to as platymeria and platycnemia, respectively. However, all the postcranial elements are consistent with a habitually upright posture and long-range bipedalism.

In 1964, Louis Leakey, Phillip Tobias, and John Napier announced the discovery at Olduvai Gorge of specimens that they believed belonged to a previously unknown species of *Homo*, which they called *H. habilis* (Leakey et al. 1964). These specimens (OH, 4, 6, 7, 8, 13, 14, and 16) were found between 1959, the same year a new genus and species, *Zinjanthropus boisei*, had been created for the famous “nutcracker man” cranium, OH 5, and 1963. The type specimen of *H. habilis*, OH 7, recovered in 1960 consists of substantial parts of both parietal bones, much of a mandible and several hand bones of a juvenile skeleton. In the next 3 years further evidence of a “non-robust” fossil hominid was unearthed in Bed I of Olduvai Gorge (OH 4 and 6, skull fragments and teeth; OH 8, an adult foot; OH 14, juvenile cranial fragments; and OH 16, the fragmented cranial vault and maxillary dentition of a young adult) as well as in Bed II (OH 13, the
incomplete skull of an adolescent). The inclusion of this group of specimens in *Homo* substantially widened the range of morphology within the genus and meant that Le Gros Clark’s (1955) diagnosis needed to be amended. In particular, in order to accommodate *H. habilis* in the genus, Leakey et al. (1964) reduced the lower end of the range of brain size to 600 cm$^3$. They claimed that other criteria, such as dexterity, an erect posture, and a bipedal gait, did not need to be changed because their interpretation of the functional capabilities of the *H. habilis* remains from Olduvai was such that the type specimen and the paratypes complied with these functional criteria (Leakey et al. 1964). Ultimately fresh evidence, and the reinterpretation of existing evidence, has led others to offer rather different functional assessments of the same material (see below).

The systematic interpretation of *Homo* was further complicated in 1972 by Richard Leakey and colleagues’ discovery of KNM-ER 1470. Recovered from the Upper Burgi Member of the Koobi Fora Formation, KNM-ER 1470 is now reliably dated to around 1.88–1.90 million years before present (Ma). Morphologically, it presents a unique mixture of a relatively large *Homo*-like neurocranium and a large, broad *Paranthropus*-like face. The presence of these two morphologies in the same cranium posed a difficulty for researchers. Which was the homoplasy—the large brain or the large face? Alone among the early commentators, Alan Walker (1976) cautioned that KNM-ER 1470 may represent a large-brained australopith. Most researchers chose the face as the site of homoplasy and argued that the large neurocranium allied the specimen with *Homo* (Leakey 1973; Rak 1987; Bilsborough and Wood 1988). As a consequence, early members of the genus *Homo* subsumed a substantially wider range of facial morphology than it did prior to the discovery of KNM-ER 1470 (Wood 1991).

In due course, additional specimens from Koobi Fora (e.g., 1590, 1802, 1813, 3732) (Wood 1991) and Olduvai (e.g., OH 62) (Johanson et al. 1987) were added to the early *Homo* hypodigm, as was fossil evidence from Members G and H of the Shungura Formation (Howell and Coppens 1976; Boaz and Howell 1977; Coppens 1980), Member 5 at Sterkfontein (Hughes and Tobias 1977; Clarke 1985), and Member 1 at Swartkrans (Clarke and Howell 1972; Grine et al. 1993, 1996; Grine and Strait 1994). This additional material subsumes a wide range of cranial morphology. For example, the endocranial volumes of the specimens range from just less than 500 cm$^3$ to around 850 cm$^3$. The mandibles also vary in size, with those from the larger individuals having robust bodies and premolar teeth with complex crowns and roots. The discovery of OH 62 was particularly significant with regard to the postcranial anatomy of *H. habilis*. Although the preservation of this specimen is poor, its skull is sufficiently well preserved to be confident that it does not belong to *Paranthropus boisei*. Thus, unless it is the first evidence from Bed I of a novel taxon, then OH 62 must belong
to *H. habilis*, the only other hominid species known from that time range at Olduvai Gorge. Although several isolated postcranial specimens from Bed I had been attributed to *H. habilis* (Leakey et al. 1964), it was subsequently pointed out that it is at least equally likely that this postcranial evidence belongs to *P. boisei* (Wood 1974). The discovery of OH 62 provided the first unequivocal postcranial evidence of *H. habilis*. It is significant therefore that OH 62 has been interpreted as having limb proportions that are at least as apelike as those of individuals attributed to *Australopithecus afarensis* (Johanson et al. 1987; Hartwig-Scherer and Martin 1991; Richmond et al. 2002).

The morphological limits of genus *Homo* were expanded once again in 2004 with the announcement of the species *H. floresiensis* (Brown et al. 2004). The specimens initially attributed to this species were recovered from deposits in the Liang Bua cave on the Indonesian island of Flores and are dated to between approximately 74 and 18 ka (Brown et al. 2004; Morwood et al. 2004). They include a well-preserved skull and partial skeleton of an adult female as well as several more fragmentary specimens (Brown et al. 2004). Since then additional upper limb evidence of the type specimen, LB1, has been recovered, together with a second adult mandible (LB6), and postcranial remains belonging to other individuals (LB4, 5, 7, 8–9) (Morwood et al. 2005). These new finds have changed the last appearance date for *H. floresiensis* from 18 to 12 ka (Morwood et al. 2005). *H. floresiensis* is a particularly significant addition to *Homo* because of its brain size. The endocranial volume of the partial associated female skeleton, LB1, was initially reported to be 380 cm$^3$ (Brown et al. 2004). Subsequently, Falk et al. (2005) increased this figure to 417 cm$^3$. Even at 417 cm$^3$, the endocranial volume of *H. floresiensis* is considerably smaller than those of the other species assigned to *Homo*. Among the latter, *H. habilis*, *H. rudolfensis*, and *H. erectus* (including *H. ergaster*) have the smallest endocranial volumes. Adult endocranial volume in *H. habilis* ranges between 509 and 674 cm$^3$ (Tobias 1991; Kappelman 1996). Only one of the specimens assigned to *H. rudolfensis* is sufficiently complete to provide an adult endocranial volume for this species. The specimen in question, KNM-ER 1470, is estimated to have an endocranial volume of 752 cm$^3$ (Kappelman 1996). Currently the upper limit of adult endocranial volume in *H. erectus* is 1,251 cm$^3$ (Rightmire 2004); the lower limit is either 775 cm$^3$ or $\sim$600 cm$^3$ depending on the taxonomic status of the D2700 cranium from Dmanisi (Vekua et al. 2002; Rightmire 2004). Thus, the assignment of the Late Pleistocene Liang Bua specimens to *Homo* greatly increases brain size variation in the genus. The body of *H. floresiensis* has been suggested to be small compared to other species of *Homo* (Lahr and Foley 2004). However, the stature estimates of 106 cm for LB1 (Brown et al. 2004) and 109 cm for LB8 (Morwood et al. 2005) are only slightly smaller than
McHenry’s (1991) stature estimate of 118 cm for the *H. habilis* partial skeleton OH 62.

Interpretations of *Homo* have also changed as a result of researchers reasessing the functional implications of the postcranial remains from Olduvai Gorge that are conventionally attributed to *H. habilis*. The type and paratypes of *H. habilis* include fossil evidence from both the forelimb (OH 7) and the hindlimb (OH 8, 10, and 35) (some have argued that OH 8 and 35 are from the same individual [Susman and Stern 1982], but an analysis of the shapes of the reciprocal joint surfaces suggests otherwise [Wood et al. 1998]). The initial assessment of the functional implications of the evidence from the leg and foot stressed the ways in which the Olduvai material resembled *H. sapiens* (Napier 1964). However, the authors of papers in which these specimens have been considered in more detail have been more cautious. For example, they have stressed that the knee was imperfectly adapted to bipedalism (Davis 1964) and that the foot may not have been from an individual capable of modern humanlike striding bipedalism (Day and Napier 1964). Functional morphological studies of the OH 8 foot have also stressed its potential for climbing and its retention of several of the features seen in the living nonhuman primates (Susman and Stern 1982; Lewis 1983, 1989; Kidd et al. 1996). Researchers have suggested that, while OH 8 possesses the articular mechanisms that convert the foot into a rigid lever during the support phase of walking (Lewis 1989), it apparently lacks some of the functional elements that are present in *H. sapiens* such as the lateral deviation of the heel and the propulsive great toe (Lewis 1972). Similarly, considerations of the OH 7 hand have suggested that earlier functional interpretations may need to be revised in the light of evidence that it displays a mosaic of features, ranging from apelike phalanges and carpus to a thumb that some have interpreted as compatible with pulp-to-pulp opposition (Susman and Creel 1979; Marzke 1997; Susman 1998).

Today, as a result of the developments outlined above, the genus *Homo* subsumes considerably more variation than it did when it was first established 250 years ago. This variation is particularly obvious in relation to cranial capacity. The adult *Homo* specimen with the largest recorded cranial capacity is the Neanderthal associated skeleton Amud 1 at an estimated 1750 cm³. At the other end of the spectrum lies the type specimen of *H. floresiensis*, Liang Bua 1, with a cranial capacity of 417 cm³. To put this in perspective, at 1750 cm³, the braincase of Amud 1 is almost 100 cm³ larger than that of the largest *H. sapiens* specimen included in one of the most comprehensive studies of human brain size published to date (Beals et al. 1984), while the braincase of Liang Bua 1 is slightly smaller than that of an average-sized adult male chimpanzee (Kappelman 1996). Variability is also conspicuous in relation to masticatory morphology. For example, the lower first molars of the *Homo* species with the largest M1s, *H. rudolfensis*,
are 32% larger mesiodistally than those of the *Homo* species with the smallest M$_1$s, *H. neanderthalensis* (Wood and Collard 1999). Likewise, average mandibular corpus width at M$_1$ in *H. rudolfensis*, the *Homo* species with the widest mandibular corpus, is 77% greater than it is in *H. sapiens*, the *Homo* species with the narrowest mandibular corpus (Wood and Collard 1999). Noteworthy variability in locomotor strategies exists as well. Most of the fossil species assigned to *Homo* are interpreted as having been obligate bipeds like *H. sapiens*. However, as noted earlier, in recent years the postcranial specimens assigned to *H. habilis* have come to be viewed by most researchers as being consistent with bipedalism combined with an ability to climb proficiently (McHenry and Coffing 2000; Wood and Richmond 2000). Thus, there would seem to be at least two distinct modes of locomotion represented within *Homo*.

### 8.3 Is genus *Homo* a “good” genus?

In view of the conspicuous variation incorporated within the hypodigm of *Homo*, it seems reasonable to ask whether it is a “good” genus as it is currently construed. Needless to say, in order to determine whether or not *Homo* is a “good” genus, there must first be agreement about what it is that genera represent.

Surprisingly, the genus concept has received relatively little attention from taxonomists. It certainly has received much less consideration than the species concept even though, as Simpson (1963 p 199) notes, “it frequently appears that the genus is a more usable and reliable unit for classification than the species.” In the paleoanthropological literature discussion of the genus as a concept has been very limited indeed. Criteria for recognizing fossil specimens of *Homo*, *Australopithecus*, and other hominid genera have been outlined on a number of occasions (Le Gros Clark 1955; Howell 1978). But these criteria have generally not been explicitly derived from a genus concept. Rather, the course of action followed by most researchers appears to have been to assign new specimens to a fossil hominid genus on the basis of a subset of the diagnostic traits that the researchers in question deem to be key and to then redefine the other traits of the genus in the light of the morphological and functional attributes of the new specimens. This is seen most clearly in relation to Leakey et al.’s (1964) proposal to recognize *H. habilis*. As noted earlier, Leakey et al. (1964) assigned the *habilis* specimens from Olduvai Gorge to *Homo* on the grounds that, according to their interpretation of the available postcranial evidence, *H. habilis* stood upright and moved around using a bipedal gait and was capable of modern humanlike dexterity. They then amended the diagnosis of *Homo* presented by Le Gros Clark (1955) to take into account the 638–674 cm$^3$ brains of the Olduvai specimens. At no point
in their paper do Leakey et al. (1964) discuss the pros and cons of assigning the Olduvai *H. habilis* specimens to *Homo* in terms of the genus as a concept.

If we wish to move beyond this *ad hoc* approach to assigning fossil hominid specimens to genera, what options are available? Currently, there appear to be six competing proposals for how genera should be defined. The first is associated with Ernst Mayr and the evolutionary systematic school of taxonomy. Mayr (1950 p 110) suggested that “a genus consists of one species, or a group of species of common ancestry, which differ in a pronounced manner from other groups of species and are separated from them by a decided morphological gap.” He went on to state that the genus “has a very distinct biological meaning. Species that are united in a given genus occupy an ecological situation which is different from that occupied by the species of another genus, or, to use the terminology of Sewall Wright, they occupy a different adaptive plateau” (Mayr 1950 p 110). Thus, according to Mayr, a genus is a group of species of common ancestry that is adaptively both homogeneous and distinctive. Mayr et al. (1953 p 50) acknowledged the phylegetic and functional evidence may be in conflict if “unrelated species acquire a superficial similarity owing to parallel adaptations to similar environments,” and in such cases they recommended that the phylegetic evidence should be given precedence. However, it is implicit in Mayr’s (1950) definition that “common ancestry” subsumes both monophyletic and paraphyletic groups.

The second concept of the genus is associated with Willi Hennig and the phylegetic systematic or cladistic school of taxonomy. In his 1966 volume “Phylogenetic Systematics,” Hennig outlined not only what has come to be known as the cladistic method of phylegetic reconstruction but also an approach to biological classification. With regard to the latter, Hennig (1966) suggested that only monophyletic groups should be accepted as valid taxa and that the ranks assigned to taxa should be based on their time of origin so that taxa arising at the same time are assigned the same rank. Hennig (1966) recognized that strict application of the time of origin criterion for delineating ranks is impractical. This is because it would involve massive reorganization of current classifications of the living world, with some groups being lumped to a much greater extent and others being split much more finely. For example, as Hennig (1966) noted, if the time of origin criterion were to be applied strictly, then the first appearance date of Mammalia is such that the class would have to be downgraded to an order, and the orders that are assigned to it, such as Primates, would have to be downgraded to tribes. Conversely, the first appearance dates of some ostracod genera are such that they would have to be elevated to the class rank. Accordingly, Hennig proposed a compromise in which different timescales are employed for different animal groups, with the timescales being selected with a view to minimizing the number of changes in rank of subgroups.
The third genus concept focuses on hybridizability as the criterion for grouping species into genera. Although it has not been widely used by systematists to date, the origins of the concept can be traced back at least as far as the middle of the nineteenth century. Flourens (1856), for example, argued that two species whose members are able to produce hybrids, such as horses and donkeys, and jackals and dogs, should be placed in the same genus. More recently, the concept has been discussed by Hubbs and Drewry (1959), Van Gelder (1977, 1978), and Dubois (1988). The most comprehensive exposition of the hybridizability-based genus concept published to date is to be found in Dubois (1988). According to this author, when two species are able to produce viable adult hybrids both species should be included in the same genus. This is the case, Dubois (1988) suggested, regardless of whether the hybrids are fertile or infertile. If the two species in question had previously been attributed to distinct genera then they should be merged together even if other criteria for separating them are valid. In other words, Dubois (1988) argued that the ability to give birth to viable adult hybrids should be the primary criterion for grouping species into genera. Dubois (1988) stressed two additional points. One is that hybridization need not take place only in the wild to be admissible as evidence of the congeneric status of two species; the results of experimental studies are also acceptable. The other point he emphasized is that the criterion of hybridizability must only be used to group species together. A negative result—i.e., one where hybridization does not occur or where the hybrid is not viable—cannot be used to place two species in different genera.

The fourth genus concept was proposed by Wood and Collard (1999). These authors suggested that a genus should be defined as a species, or monophylum, whose members occupy a single adaptive zone. This definition is essentially a revised version of Mayr’s (1950) concept. It differs from the latter in that it excludes paraphyletic taxa. It also differs from Mayr’s (1950) concept in that it does not require the adaptive zone to be unique or distinct (contra Leakey et al. 2001; Cela-Conde and Altaba 2002). Rather, it simply requires the adaptive zone to be consistent and coherent across the species in the putative genus. That is, in contrast to Mayr (1950), Wood and Collard (1999) allowed for the possibility that species assigned to different genera will occupy the same adaptive zone but prevent species in the same genus from occupying different adaptive zones. Wood and Collard (1999) suggested two criteria for assessing whether or not a group of species has been correctly assigned to a genus. First, the species should belong to the same monophyletic group as the type species of that genus. Second, the adaptive strategy of the species should be closer to the adaptive strategy of the type species of the genus in which it is included than to the type species of any other genus.

The fifth approach to recognizing genera was outlined by Watson et al. (2001). These authors suggested that species should be grouped into genera on
the basis of genetic distance. Specifically, Watson et al. (2001) argued that, if the genetic distance between a pair of species is the same as or less than the genetic distance that is typical for congeneric pairs of species in other animal groups, then the species in question should be assigned to the same genus.

The sixth approach to the genus concept was outlined by Cela-Conde and Altaba (2002). Their concept is similar to the one advocated by Wood and Collard (1999) in that it holds that a genus should be monophyletic and uses inferences about adaptation to determine which monophyla should be designated genera. However, it differs from Wood and Collard’s (1999) concept in that it allows for the species assigned to a genus to occupy more than one adaptive zone. Specifically, Cela-Conde and Altaba (2002) suggested that one species in each genus should be designated as the *species germinalis*. This species is the one that is considered to have given rise to the other species in the genus. Because of its ancestral status, the *species germinalis* is allowed to occupy a different adaptive zone from the other species in the genus.

Several of these concepts have been applied to genus *Homo* in recent years. Hennig’s (1966) approach to delineating supraspecific taxa has been applied to *Homo* in a series of papers and books produced by a team of researchers led by Morris Goodman (Goodman et al. 1998; Goodman et al. 2001; Wildman et al. 2003). Goodman et al. (1998) averred that, at least among the primates, monophyla should be deemed to be genera if they originated 11–7 Ma. Goodman et al. (1998) selected this criterion on the basis of local molecular clock analyses, which suggested that the majority of extant primate genera arose between 11 and 7 Ma. Subsequently, Wildman et al. (2003) further justified the selection of an origination date of 11–7 Ma as the criterion for recognizing primate monophyla as genera on the grounds that the majority of genera in other mammalian orders arose between 11 and 7 Ma. Significantly for present purposes, as Goodman and coworkers have noted on a number of occasions (Goodman et al. 1998, 2001; Wildman et al. 2003), their definition of the genus implies that *Homo* should be broadened to include not only the australopiths and other early hominids but also chimpanzees and bonobos which are conventionally assigned to the genus *Pan*. This is because molecular clock studies suggest that humans, chimpanzees, and bonobos last shared a common ancestor around 6 Ma and the nomen *Homo* has priority over the nomen *Pan*, the former being proposed by Linnaeus in 1758, as noted earlier, and the latter by Oken in 1816. Subsequently, one of the authors of the Goodman et al. (1998) study, Colin Groves, proposed at least two other time depth-based criteria for recognizing monophyla as genera (Groves 2001a, b; Cameron and Groves 2004). For example, in his widely cited monographic treatment of primate taxonomy, Groves reviewed evidence pertaining to the chronological origin of genera in several mammalian families, including Ursidae, Canidae, Elephantidae, Rhinocerotidae, Hippopotamidae, and Bovidae, and concluded on the basis of
this evidence that an origin time of between 7 and 4 Ma should be used as the criterion for delineating extant mammalian genera (Groves 2001a). This led him to retain *Homo* and *Pan* as separate genera in contrast to Goodman et al. (1998). More recently, Groves proposed that primate monophyla should be recognized as genera if they originated between 6 and 4 Ma (Cameron and Groves 2004). The corollary of this, he suggested, is that all extinct hominid genera and perhaps also the chimpanzee genus, *Pan*, should be assigned to *Homo*.

In the paper in which Wood and Collard outlined their genus concept (Wood and Collard 1999), they applied the criteria derived from it to the species that most researchers then assigned to genus *Homo*, namely *H. erectus*, *H. ergaster*, *H. habilis*, *H. heidelbergensis*, *H. neanderthalensis*, *H. rudolfensis*, and *H. sapiens*. They examined a range of phylogenetic and functional evidence in order to determine whether or not the fossil species assigned to *Homo* form a monophylum with *H. sapiens* and also share its adaptive strategy. They found that the only fossil *Homo* species that form a robust clade with *H. sapiens* are *H. neanderthalensis*, *H. heidelbergensis*, *H. erectus*, and *H. ergaster*. They also found that when evidence about body size, body shape, relative brain size, and development is combined with inferences about locomotion and diet, these species are the only *Homo* taxa whose adaptations are closer to those of *H. sapiens* than they are to those of *A. africanus*, the type species of *Australopithecus*. The phylogenetic relationships of *H. habilis* and *H. rudolfensis* were found to be equivocal, and the available evidence regarding the adaptive strategies of *H. habilis* and *H. rudolfensis* was interpreted as indicating that they were at least as similar, and probably more similar, to the australopiths than they are to *H. sapiens*. Wood and Collard (1999) concluded from this that a genus *Homo* that includes them is not a “good” genus and that *H. habilis* and *H. rudolfensis* should be removed from genus *Homo* and placed in *Australopithecus* until such time as their phylogenetic relationships are clarified. Recently, Cameron has employed the concept proposed by Wood and Collard (1999) but reached different conclusions regarding the fossil species that should be assigned or excluded to *Homo* (Cameron and Groves 2004). Most notably for present purposes, he argued that the *H. habilis* hypodigm should be retained in *Homo* and suggested that the *H. rudolfensis* hypodigm should be removed from *Homo* and assigned to *Kenyanthropus*. The latter proposal is based on cladistic analyses that link the *H. rudolfensis* hypodigm with *Kenyanthropus platyops* (Cameron and Groves 2004). Cameron does not explain how retaining the *H. habilis* hypodigm within *Homo* is consistent with the notion that a genus should be a species or monophylum whose members occupy a single adaptive zone. However, the implication is that he does not accept Wood and Collard’s (1999) contention that the adaptive strategy of *H. habilis* was more similar to that of *A. africanus* than to the adaptive strategy of *H. sapiens*. 
In the 2001 paper in which Watson and colleagues outlined their genetic distance-based concept of the genus they also applied the concept to previously published mtDNA sequence and DNA hybridization data for humans, chimpanzees, gorillas, and a range of other mammalian groups with a view to classifying the living hominoids (Watson et al. 2001). They found that the genetic distances between chimpanzees and humans are equivalent to the distances between many mammalian species within the same genus. They also found that the genetic distances between gorillas and chimpanzees, and between gorillas and humans, are similar to the distances between congeneric mammalian species. These observations, Watson et al. (2001) suggested, indicate that the genus *Homo* should be expanded to include chimpanzees and gorillas as well as humans. Watson et al.’s (2001) approach has also been applied to DNA distance data for humans and chimpanzees by Curnoe and Thorne (2003). These authors also concluded that the human and chimpanzee genomes are sufficiently similar for the species to be considered congeneric. Accordingly, they recommended transferring chimpanzees to genus *Homo*.

In 2002, Cela-Conde and Altaba revised the taxonomy for the hominids proposed by Wood and Collard (1999) in the light of their *species germinalis* concept and fossil specimens recovered in the intervening period (Cela-Conde and Altaba 2002). Most significantly for present purposes, Cela-Conde and Altaba (2002) agreed with Wood and Collard (1999) that *H. rudolfensis* should be removed from genus *Homo* but disagreed with them regarding the generic attribution of *H. habilis*. They suggested that *H. rudolfensis* should be transferred to the genus *Kenyanthropus*, which had been erected in 2001 to accommodate the newly discovered species *K. platyops* and that *H. habilis* should be included in *Homo* as the *species germinalis* of the genus. In 2003, Cela-Conde and Ayala revised the taxonomy proposed by Cela-Conde and Altaba (2002) (Cela-Conde and Ayala 2003). They argued not only that *H. habilis* and *H. rudolfensis* should be included in *Homo* but also that the hypodigm of *K. platyops* should be transferred to *Homo* as the *species germinalis* of the genus. Cela-Conde and Ayala (2003) included the *H. habilis* and *H. rudolfensis* hypodigms in *Homo* on the grounds that they share the morphological traits that Leakey et al. (1964) suggested define *Homo*. The inclusion of the *K. platyops* hypodigm in *Homo* is justified, Cela-Conde and Ayala (2003) argued, because it is similar to *H. habilis* and especially *H. rudolfensis* in certain features of its face and dentition. Cela-Conde and Ayala (2003) designated *platyops* as the *species germinalis* of *Homo* because it lacks “the more advanced features of *Homo* that appear with *Homo erectus* and *Homo ergaster*” (p 7686).

Table 8.1 presents an illustrative comparison of the definitions put forward by Goodman et al. (1998), Wood and Collard (1999), Watson et al.
**Table 8.1**

Illustrative comparison of the impact of four recent proposals to revise genus *Homo* on the taxonomy of hominids and the African apes. A conventional splitter’s taxonomy (CST) taken from Stanford et al. (2005) is used as the baseline for comparison.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Gorilla gorilla</td>
<td>Gorilla gorilla</td>
<td>Gorilla gorilla</td>
<td>Homo gorilla</td>
<td>Gorilla gorilla</td>
</tr>
<tr>
<td>Pan paniscus</td>
<td>Homo paniscus</td>
<td>Pan paniscus</td>
<td>Homo paniscus</td>
<td>Pan paniscus</td>
</tr>
<tr>
<td>Pan troglodytes</td>
<td>Homo troglodytes</td>
<td>Pan troglodytes</td>
<td>Homo troglodytes</td>
<td>Pan troglodytes</td>
</tr>
<tr>
<td>Ardipithecus ramidus</td>
<td>Homo ramidus</td>
<td>Ardipithecus ramidus</td>
<td>Homo ramidus</td>
<td>Ardipithecus ramidus</td>
</tr>
<tr>
<td>Australopithecus afarensis</td>
<td>Homo afarensis</td>
<td>Australopithecus afarensis</td>
<td>Homo afarensis</td>
<td>Australopithecus afarensis</td>
</tr>
<tr>
<td>Australopithecus africanus</td>
<td>Homo africanus</td>
<td>Australopithecus africanus</td>
<td>Homo africanus</td>
<td>Australopithecus africanus</td>
</tr>
<tr>
<td>Australopithecus anamensis</td>
<td>Homo anamensis</td>
<td>Australopithecus anamensis</td>
<td>Homo anamensis</td>
<td>Australopithecus anamensis</td>
</tr>
<tr>
<td>Australopithecus bahrelghazali</td>
<td>Homo bahrelghazali</td>
<td>Australopithecus bahrelghazali</td>
<td>Homo bahrelghazali</td>
<td>Australopithecus bahrelghazali</td>
</tr>
<tr>
<td>Australopithecus garhi</td>
<td>Homo garhi</td>
<td>Australopithecus garhi</td>
<td>Homo garhi</td>
<td>Australopithecus garhi</td>
</tr>
<tr>
<td>Homo erectus</td>
<td>Homo erectus</td>
<td>Homo erectus</td>
<td>Homo erectus</td>
<td>Homo erectus</td>
</tr>
<tr>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Homo habilis</td>
<td>Homo habilis</td>
<td>Australopithecus habilis</td>
<td>Homo habilis</td>
<td>Homo habilis</td>
</tr>
<tr>
<td>Homo heidelbergensis</td>
<td>Homo heidelbergensis</td>
<td>Homo heidelbergensis</td>
<td>Homo heidelbergensis</td>
<td>Homo heidelbergensis</td>
</tr>
<tr>
<td>Homo neanderthalensis</td>
<td>Homo neanderthalensis</td>
<td>Homo neanderthalensis</td>
<td>Homo neanderthalensis</td>
<td>Homo neanderthalensis</td>
</tr>
<tr>
<td>Homo rudolfensis</td>
<td>Homo rudolfensis</td>
<td>Australopithecus rudolfensis</td>
<td>Homo rudolfensis</td>
<td>Homo rudolfensis</td>
</tr>
<tr>
<td>Homo sapiens</td>
<td>Homo sapiens</td>
<td>Homo sapiens</td>
<td>Homo sapiens</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>Kenyanthropus platyops</td>
<td>Homo platyops</td>
<td>Kenyanthropus platyops</td>
<td>Homo platyops</td>
<td>Homo platyops</td>
</tr>
<tr>
<td>Orrorin tugenensis</td>
<td>Homo tugenensis</td>
<td>Orrorin tugenensis</td>
<td>Homo tugenensis</td>
<td>Orrorin tugenensis</td>
</tr>
<tr>
<td>Paranthropus aethiopicus</td>
<td>Homo aethiopicus</td>
<td>Paranthropus aethiopicus</td>
<td>Homo aethiopicus</td>
<td>Paranthropus aethiopicus</td>
</tr>
<tr>
<td>Paranthropus boisei</td>
<td>Homo boisei</td>
<td>Paranthropus boisei</td>
<td>Homo boisei</td>
<td>Paranthropus boisei</td>
</tr>
<tr>
<td>Paranthropus robustus</td>
<td>Homo robustus</td>
<td>Paranthropus robustus</td>
<td>Homo robustus</td>
<td>Paranthropus robustus</td>
</tr>
<tr>
<td>Sahelanthropis tchadensis</td>
<td>Homo tchadensis</td>
<td>Sahelanthropis tchadensis</td>
<td>Homo tchadensis</td>
<td>Sahelanthropis tchadensis</td>
</tr>
</tbody>
</table>
(2001), and Cela-Conde and Ayala (2003). The first column in the table lists the names of the genera and species recognized in a typical taxonomy covering the hominids and the extant African apes (Stanford et al. 2005). The other columns in the table record the names of the species and genera that would be recognized among the hominids and African apes if the conventional taxonomy were revised in line with the proposals of Goodman et al. (1998), Wood and Collard (1999), Watson et al. (2001), and Cela-Conde and Ayala (2003). The table shows that the four definitions have different implications for not only the composition of genus Homo but also the taxonomy of hominids and extant African apes. The conventional taxonomy recognizes 21 species and assigns these to 9 genera. Revising the conventional taxonomy in line with Wood and Collard’s (1999) and Cela-Conde and Ayala’s (2003) proposals results in relatively few changes. Wood and Collard’s (1999) scheme leads to species being moved between genera but no reduction in the number of genera, while the one put forward by Cela-Conde and Ayala (2003) requires species to be moved between genera and the elimination of a single genus, *Kenyanthropus*. Goodman et al.’s (1998) and Watson et al.’s (2001) proposals have more radical implications. If the conventional taxonomy were to be reorganized in line with Goodman et al.’s (1998) definition of *Homo*, then the 21 species would be assigned to just two genera, *Homo* and *Gorilla*. *Homo* would have 20 species assigned to it and *Gorilla* a single species. If the conventional taxonomy were reorganized in line with the definition of *Homo* offered by Watson et al. (2003), then the 21 species would be assigned to a single genus. Thus, some proposals to redefine *Homo* have little or no impact on the current consensus regarding generic diversity among the hominids and African apes, while others have a major impact.

The definitions of genus *Homo* presented by Goodman et al. (1998), Wood and Collard (1999), Watson et al. (2001), Cela-Conde and Altaba (2002), and Cela-Conde and Ayala (2003) have other implications for how *Homo* is interpreted. One of the most obvious is the time of its origin. Conventional taxonomies such as the one outlined in Table 8.1 suggest that the genus arose in the Late Pliocene, since the first appearance dates of *H. habilis* and *H. rudolfensis* are 2.3 and 2.5 Ma, respectively. Collard and Wood’s (1999) definition, which excludes *H. habilis* and *H. rudolfensis* from the genus, implies that *Homo* appeared about 2 Ma, shortly after the start of the Pleistocene. In contrast, the other three definitions push back the date of origin. Cela-Conde and Ayala’s (2003) definition implies that the genus first appeared around 3.5 Ma, while the definition put forward by Goodman et al. (1998) implies that the genus originated about 7–8 Ma. Watson et al.’s (2001) definition pushes back the origin of genus *Homo* by at least another million years, since the lineage leading to gorillas
and the lineage leading to chimpanzees and humans are estimated to have split around 9 Ma (Ruvolo 1997). Another aspect of the genus that varies considerably depending on the definition employed is its mode of locomotion. Conventional taxonomies incorporate at least two forms of locomotion, facultative bipedalism and obligate bipedalism. The extinct hominid species that Wood and Collard (1999) assign to *Homo* are all reconstructed as being obligate bipeds, while Goodman et al.’s (1998) and Watson et al.’s (2001) definitions incorporate obligate bipeds, facultative bipeds, and knuckle-walkers within genus *Homo*. Most of the other adaptive characteristics that are of interest to paleoanthropologists, such as the size of the masticatory system, brain size, and developmental schedule, are affected in a similar manner.

Given that the genus concepts evidently have markedly different implications for the composition, and therefore the interpretation, of *Homo*, which of them is to be preferred? In our view, the approach to delineating genera proposed by Watson et al. (2001) is not convincing. The notion that genetic distances among congeneric species in one animal group should be used as a criterion to cluster species into genera in another animal group is problematic. First, given that there does not seem to be a straightforward relationship between genetic distance and morphological distance among living taxa (Lambert and Paterson 1993), it is doubtful that Watson et al.’s (2001) suggestions can be applied to fossil taxa. Second, even if it were possible to obtain reliable estimates of the genetic distances among pairs of fossil species, there is no reason to believe that interspecific genetic distances are distributed in such a way as to justify designating any particular distance or range of distances as the criterion for clustering species into genera. This course of action might be appropriate if interspecific genetic distances were discontinuously distributed or if there were sound theoretical reasons why genera should correspond to a given genetic distance. However, neither of these conditions appears to be the case. The available evidence suggests that genetic distances are more or less continuously distributed (Lambert and Paterson 1993), and Watson et al. (2001) do not provide any theoretical justification for the approach they advocate. Accordingly, there is no reason to prefer one particular genetic distance or range of distances as the criterion for delineating genera over any other genetic distance or range of distances.

We are also skeptical about the utility of the hybridizability-based concept for assigning species to genus *Homo*. Given that many of the relevant species are known only from fossilized hard tissues, application of this concept requires skeletal variation among hominids to be a reliable proxy for hybridizability. This assumption is only valid if skeletal variation has been found to accurately predict the limits of hybridizability in a range of appropriate model taxa such as
the living primates. Although to the best of our knowledge the relationship between skeletal variation and hybridizability has not been investigated in any relevant animal group to date, there is good reason to doubt that the relationship between these parameters in primates is such that the hybridizability-based concept can be used to assign species to genus *Homo*. In the last 20 years, a number of studies have demonstrated that there is considerable overlap between intraspecific and interspecific skeletal variation in living primates (Tattersall 1986; Kimbel 1991; Aiello et al. 2000). Given that species status has been conferred on most extant primate species on the basis of failure to produce fertile offspring in the wild, this overlap suggests that skeletal morphology is a poor guide to reproductive biology in primates. Accordingly, it seems unlikely that skeletal morphology can be used to assign fossil hominid species to *Homo* on the basis of their likely ability to produce viable hybrids.

Hennig’s (1966) approach to delineating genera and other supraspecific taxa has the advantage that it is relatively easy to implement. Another advantage of Hennig’s approach for paleoanthropology, at least in the way it has been implemented by Goodman et al. (1998), is that the genera it produces can be expected to be relatively stable. Given that, as discussed earlier, the available molecular data suggest that the split between the human and chimpanzee lineages occurred no earlier than 8 Ma, Goodman et al.’s (1998) proposal that monophyla should be recognized as genera if they originated between 11 and 7 Ma means that new fossil hominid finds and fresh phylogenetic analyses are unlikely to require the creation of new genera.

However, the time-based approach also has a number of shortcomings. One of these concerns the manner in which the time ranges that correspond to different taxonomic ranks are chosen. As noted earlier, the approach that is most defensible on theoretical grounds—strict application of the time of origin criterion across all groups—was considered to be impractical even by Hennig. Unfortunately, the alternative approach proposed by Goodman et al. (1998) and Groves (2001a)—assigning taxa in one group of organisms (e.g., primates) to ranks on the basis of the age of origin of taxa within another group of organisms (e.g., bears)—is problematic. One problem is that the approach is sensitive to the choice of comparator groups. For example, Goodman et al. (1998) review one set of first appearance dates and conclude that monophyla should be recognized as genera if they originate between 11 and 7 Ma, while Groves (2001a) reviews another set of first appearance dates and concludes that the relevant time span should be 6–4 Ma. It is difficult to see how this can be avoided given that phylogenetic relationships are relative phenomena. How do we defend a given degree of relatedness as the criterion for deciding which taxa to include in our comparator group and which to exclude?
A second and perhaps even more profound problem is that the approach proposed by Goodman et al. (1998) and Groves (2001a) is not internally consistent. Again, the problem lies with the comparator taxa. Given that, as we noted earlier, strict application of the time of origin across all groups is impractical, at least one of the comparator taxa must be defined in relation to a criterion other than time such as adaptive coherence. For example, as noted earlier, Groves (2001a) assigns primate species to genera in the light of the first appearance dates of genera in Ursidae, Canidae, Elephantidae, Rhinocerotidae, and Hippopotamidae, but it is evident from the references he cites in relation to the latter that they have been defined on the basis of anatomical evidence. Thus, the approach advocated by Goodman et al. (1998) and Groves (2001a) essentially entails reorganizing the taxonomy of one group of organisms on the basis of prior taxonomic analyses of another group of organisms that employed a different approach to delineating taxa. We recognize that, in the absence of a strict application of the time of origin across all groups, this is unavoidable. However, this does not prevent it from being a major flaw. It means that the approach is not only contradictory (time of origin is the preferred criterion for assigning taxa to ranks except in the case of the comparator taxa, which are defined in relation to some other criterion) but also effectively requires paleoanthropologists to subjugate their own taxonomic philosophies in favor of those used by researchers working on other groups of organisms.

The differences between the remaining approaches are subtle. To reiterate, for Mayr (1950) a genus is a species or group of species of common descent that occupies an ecological situation that is different from those occupied by the species of another genus, while for Wood and Collard (1999) a genus is a species, or monophylum, whose members occupy a single adaptive zone. Cela-Conde and Altaba’s (2002) concept is similar to the one proposed by Wood and Collard (1999) in that it holds that the species assigned to a genus should be monophyletic and uses inferences about adaptation to determine which monophyla should be designated genera. However, it differs from Wood and Collard’s (1999) concept in that one species, the species germinalis, is allowed to occupy a different adaptive zone from the other species in the genus. Thus, the approaches differ regarding whether phylony should be given priority over adaptation or vice versa. Mayr’s (1950) approach prioritizes species’ adaptive characteristics over their phylogenetic relationships, Wood and Collard’s (1999) approach prioritizes species’ phylogenetic relationships but also takes into account their adaptive characteristics, and Cela-Conde and Altaba’s (2002) approach prioritizes species’ phylogenetic relationships over their adaptive characteristics. One important consequence of these differences is that Mayr’s (1950) approach allows genera to be either monophyletic or paraphyletic, whereas the approaches favored by
Wood and Collard (1999) and Cela-Conde and Altaba (2002) dictate that genera have to be monophyletic. Another important consequence is that Wood and Collard’s (1999) approach makes allowance for the possibility that species in different genera will occupy the same adaptive zone, whereas Mayr’s (1950) approach demands that species assigned to different genera must have different adaptive strategies. In contrast to both Mayr’s (1950) approach and the one put forward by Wood and Collard (1999), Cela-Conde and Altaba’s (2002) approach anticipates that the species assigned to a genus may subsume two adaptive strategies—the ancestral adaptive strategy, which will be displayed by the *species germinalis*, and the derived adaptive strategy, which will be exhibited by the remaining species.

Choosing between the genus concepts proposed by Mayr (1950), Wood and Collard (1999), and Cela-Conde and Altaba (2002) is not straightforward. It is especially difficult for paleoanthropologists, given that they often need to classify taxa based on a few specimens, and occasionally just a single specimen. The dilemma with which they are confronted was outlined particularly clearly by Alan Walker in a paper titled “Remains attributable to *Australopithecus* in the East Rudolf succession,” published in 1976. Walker pointed out that while classifying fossil hominids on the basis of their adaptive characteristics is problematic, so too is classifying them on the basis of their phylogenetic relationships. The former is problematic because, if the approach is followed to its logical conclusion, there will come a point where one generation is in one taxon and the next in another. The latter is problematic because, if its logic is followed, there will come a point where part of a single population will be in one taxon and another part of the same population will be in a different taxon. None of three remaining genus concepts avoids this dilemma. Ultimately, they all force researchers to choose to err in one direction or the other. Indeed, given that evolution involves both descent and modification, it is difficult to see how it could be otherwise. Thus, it is not easy to choose between the three concepts on theoretical grounds.

Choosing between the concepts on practical grounds is also difficult. Given that one of the main purposes of a biological classification is to communicate information about taxa (Harrison 1993), the key practical issues would seem to be stability, minimizing the number of genera, and the transparency and utility of the criteria used for delineating genera. Mayr’s (1950) approach can be expected to be more stable than the approaches put forward by Wood and Collard (1999) and Cela-Conde and Altaba (2002). This is because Mayr’s (1950) approach allows genera to be either monophyletic or paraphyletic, whereas the approaches of Wood and Collard (1999) and Cela-Conde and Altaba (2002) require genera to be monophyletic. The ability of Mayr’s (1950) approach to take into account paraphyletic taxa means that the genera it produces are less likely to need revising.
in the light of new fossil finds or fresh phylogenetic analyses than the genera produced by Wood and Collard’s (1999) and Cela-Conde and Altaba’s (2002) approaches. The ability of Mayr’s (1950) approach to take into account paraphyletic taxa also means that it can be expected to result in fewer genera than the approaches of Wood and Collard (1999) and Cela-Conde and Altaba (2002). Cela-Conde and Altaba’s (2002) approach can also be expected to result in fewer genera than the approach put forward by Wood and Collard (1999) because the species that Cela-Conde and Altaba (2002) designate as the \textit{species germinalis} would be assigned to a new genus in Wood and Collard’s (1999) approach. Thus, in terms of stability and minimizing genera, Mayr’s (1950) approach is to be preferred over the approaches put forward by Cela-Conde and Altaba (2002), which in turn is to be preferred over Wood and Collard’s (1999) approach.

However, Wood and Collard’s (1999) approach has the advantage with respect to the explicitness of the criteria for delineating genera. As noted earlier, Wood and Collard (1999) suggest that the fossil species assigned to genus \textit{Homo} should be (1) more closely related to the type species of the genus, \textit{H. sapiens}, than to the type species of any other genus and (2) more similar to \textit{H. sapiens} than to the type species of any other genus in terms of key adaptive parameters. The adaptive variables they proposed should be examined are body mass, body shape, locomotion, size of the teeth and jaws, relative brain size, and developmental schedule, all of which are capable of being inferred with a reasonable degree of reliability from the fossil record. In contrast, neither Mayr (1950) nor Cela-Conde and Altaba (2002) provided workable criteria for delineating genera. Mayr (1950) suggested that the species assigned to a genus should be separated from other groups of species by a “decided morphological gap” and also occupy a “different ecological situation.” But he did not specify what constitutes a decided morphological gap or a different ecological situation. Cela-Conde and Altaba (2002) proposed that a genus should be a monophylum whose members are adaptively distinct apart from the \textit{species germinalis}, which is allowed to have the same adaptive strategy as another genus. But they did not provide criteria for determining that a group of species is adaptively distinct from another group of species. They also did not provide criteria for identifying the \textit{species germinalis}. Needless to say, the lack of adequate criteria for delineating genera makes it difficult to implement the approaches put forward by Mayr (1950) and Cela-Conde and Altaba (2002). It also makes the resulting taxonomies difficult to defend. For example, Cela-Conde and Ayala (2003) revise Cela-Conde and Altaba’s (2002) taxonomy without recourse to analysis. Accordingly, for the time being (i.e., until Mayr’s [1950] and Cela-Conde and Altaba’s [2002] approaches are operationalized satisfactorily) our preference is to use the approach outlined by Wood and Collard (1999).
8.4 Updating Wood and Collard’s (1999) review of genus Homo

With the last point of the foregoing section in mind, the remainder of this chapter is devoted to updating Wood and Collard’s (1999) review of genus Homo in the light of developments since their study appeared. To reiterate, Wood and Collard (1999) applied their genus concept to the species that most researchers assigned to genus Homo in the late 1990s, namely H. erectus, H. ergaster, H. habilis, H. heidelbergensis, H. neanderthalensis, H. rudolfensis, and H. sapiens. They examined a range of phylogenetic and functional evidence in order to determine whether or not the fossil species assigned to Homo form a monophylum with H. sapiens and also share its adaptive strategy. They suggested that the only fossil Homo species that form a robust clade with H. sapiens are H. neanderthalensis, H. heidelbergensis, H. erectus, and H. ergaster. They also found that when evidence about body size, body shape, relative brain size and development is combined with inferences about locomotion and diet, these are the only Homo species whose adaptations are closer to those of H. sapiens than they are to A. africanus, the type species of Australopithecus, or P. robustus, the type species of Paranthropus. Wood and Collard (1999) found the phylogenetic relationships of H. habilis and H. rudolfensis to be equivocal and interpreted the available evidence regarding the adaptive strategies of H. habilis and H. rudolfensis as indicating that they are more similar to A. africanus than they are to H. sapiens. Wood and Collard (1999) concluded from this that Homo is not a “good” genus and that H. habilis and H. rudolfensis should be removed from genus Homo and placed in Australopithecus until such time as their phylogenetic relationships are clarified.

Wood and Collard’s (1999) conclusions regarding the phylogenetic relationships of H. erectus, H. ergaster, H. habilis, H. heidelbergensis, H. neanderthalensis, H. rudolfensis, and H. sapiens were based on the results of the six studies that had, at the time of writing, adequately tested the monophyly of Homo (Chamberlain 1987; Chamberlain and Wood 1987; Wood 1991, 1992; Lieberman et al. 1996; Strait et al. 1997) as well as reanalyses of the datasets used in three of the studies (Wood 1991, 1992; Strait et al. 1997). Since Wood and Collard’s (1999) study appeared, a further three cladistic studies have included representatives of sufficient fossil Homo species to potentially shed some light on the phylogenetic status of the genus (Curnoe 2001; Cameron and Groves 2004; Strait and Grine 2004). Curnoe’s (2001) study focused on the phylogenetic relationships of three specimens from Southern Africa, SK 847, SK15, and Stw 53, all of which
are usually considered to represent early *Homo*. Curnoe’s analysis employed 47 cranial characters recorded on SK 847, SK15, Stw 53 plus specimens assigned to *Australopithecus afarensis*, *A. africanus*, *H. erectus*, *H. habilis*, *H. rudolfensis*, *Paranthropus aethiopicus*, *P. robustus*, and *P. boisei*. He employed the data set in six analyses in which methodological choices were varied in order to avoid bias. *Pan troglodytes* was used as the outgroup in all the analyses. The results of Curnoe’s analyses do not support the hypothesis that *Homo* is a monophylum. The most parsimonious and consensus cladograms presented by Curnoe suggest that *H. habilis*, *H. erectus*, and the three Southern African early *Homo* specimens form a clade to the exclusion of the other taxa in the sample. However, *H. rudolfensis* is not linked exclusively to the other *Homo* taxa in any of the cladograms presented by Curnoe. In two of them (A and D in Curnoe’s Figure 8.1), it is the sister taxon of a clade comprising *A. africanus* and the other *Homo* taxa. In another two (B and C in Curnoe’s Figure 8.1), it is part of a large polyclade that also contains *A. africanus*. In the remaining cladogram (E in Curnoe’s Figure 8.1), *H. rudolfensis* forms a clade with *P. boisei* and *P. robustus*.

Cameron and Groves (2004) examined the phylogenetic relationships of 14 hominid species, including *H. ergaster*, *H. habilis*, *H. rudolfensis*, and *H. sapiens*. Two sets of analyses were carried out. One employed 92 characters; the other utilized only the characters that were present in *Sahelanthropus tchadensis* or *K. platyops*, of which there were 52. In both sets of analyses, three Miocene ape species were employed as outgroups. A parsimony analysis of the 92 characters returned 8 equally parsimonious trees. The consensus of these clustered *H. ergaster*, *H. habilis*, and *H. sapiens* in a clade to the exclusion of the other taxa and grouped *H. rudolfensis* in a clade with *K. platyops*. A bootstrap analysis of the 92 characters supported a sister group relationship between *H. ergaster* and *H. sapiens* but was unable to resolve the relationships of the other *Homo* taxa at the 70% level of support that is commonly used to identify statistically significant clades in such analyses (Hillis and Bull 1993). *H. habilis* and *H. rudolfensis* formed a multichotomy with *K. platyops*, *A. africanus*, a clade comprising *P. aethiopicus*, *P. boisei*, and *P. robustus*, and the aforementioned (*H. ergaster*, *H. sapiens*) clade. The results of the analyses based on 52 characters were similar. Twenty equally parsimonious cladograms were returned by a parsimony analysis. The consensus of these grouped *H. ergaster*, *H. habilis*, and *H. sapiens* in a clade to the exclusion of the other taxa and clustered *H. rudolfensis* in a clade with *K. platyops*. A bootstrap analysis supported a sister group relationship between *H. ergaster* and *H. sapiens* but was unable to resolve the relationships of the other *Homo* taxa at the 70% level. Thus, neither set of analyses supported the hypothesis that
the fossil species assigned to *Homo* form a monophyletic group with *H. sapiens*. They suggest that *H. ergaster* is more closely related to *H. sapiens* than to any other fossil hominid species but are equivocal regarding the relationships of *H. habilis* and *H. rudolfensis*.

Strait and Grine (2004) carried out a series of maximum parsimony and bootstrap analyses to examine the relationships of several hominid species including *H. ergaster*, *H. habilis*, *H. rudolfensis*, and *H. sapiens*. Their dataset comprised 109 qualitative craniodental characters and 89 craniometric characters recorded on 14 hominid species plus 7 extant nonhuman primate taxa. The consensus of the most parsimonious cladograms obtained by Strait and Grine (2004) suggests that *H. ergaster*, *H. habilis*, *H. rudolfensis*, and *H. sapiens* form a clade to the exclusion of the other species in the sample. Within the (*H. ergaster*, *H. habilis*, *H. rudolfensis*, *H. sapiens*) clade, *H. ergaster* and *H. sapiens* form a clade to the exclusion of *H. habilis* and *H. rudolfensis*. The relationships among the (*H. ergaster*, *H. sapiens*) clade, *H. habilis*, and *H. rudolfensis* are unresolved. The results of Strait and Grine’s (2004) bootstrap analyses were variable with regard to the relationships of the *Homo* species. When all the characters and taxa were analyzed together, a (*H. ergaster*, *H. sapiens*) clade was supported by 86% of the bootstrap replicates, but the relationships of the other *Homo* taxa were not resolved at the 70% level. A similar result was obtained when all the taxa but only the 109 qualitative characters were included. When *K. platyops* was dropped from the all-characters analysis, both a (*H. ergaster*, *H. sapiens*) clade and a (*H. ergaster*, *H. habilis*, *H. rudolfensis*, *H. sapiens*) clade were supported by more than 70% of the replicates. Again, a similar result was obtained when only the qualitative characters were included. Thus, Strait and Grine’s (2004) parsimony analyses support the hypothesis that *Homo* is a monophylum, but this hypothesis is not consistently supported by their bootstrap analyses.

The studies of Curnoe (2001), Cameron and Groves (2004), and Strait and Grine (2004) are consistent with Wood and Collard’s (1999) study in supporting the hypothesis that *H. ergaster* is more closely related to *H. sapiens* than to the type species of any other fossil hominid genus. The results of the studies are more ambiguous with respect to the phylogenetic relationships of *H. rudolfensis* and *H. habilis*. *H. rudolfensis* clustered exclusively with the other *Homo* species in Strait and Grine’s (2004) parsimony analyses and also in some of their bootstrap analyses, but it did not cluster exclusively with the other *Homo* species in the parsimony and bootstrap analyses reported in Curnoe’s (2001) and Cameron and Groves’ (2004) studies. The results of the parsimony analyses carried out by Curnoe (2001), Cameron and Groves (2004), and Strait and Grine (2004) offer support for the hypothesis that *H. habilis* is a member of the *Homo* clade. However, the bootstrap analyses carried out by Cameron and Groves (2004)
and some of the bootstrap analyses conducted by Strait and Grine (2004) failed to support a link between *H. habilis* and later *Homo* species at the 70% level. This suggests that little confidence can be placed in this hypothesis. On balance, the results of the three new studies are consistent with those of Wood and Collard (1999) with regard to the relationships of the fossil *Homo* species. 

In the last 7 years, there have been developments in two areas that impact Wood and Collard’s (1999) conclusions regarding the adaptive strategies of *H. erectus*, *H. ergaster*, *H. habilis*, *H. heidelbergensis*, *H. neanderthalensis*, *H. rudolfensis*, and *H. sapiens*. One of these concerns the life history strategies of the fossil species. The period of maturation of *H. sapiens* is nearly twice as long as those of the *G. gorilla* and *P. troglodytes* (Dean et al. 1986; Smith 1994). This extended ontogeny has been linked with the transmission of the numerous additional learned behaviors that modern humans exhibit compared to the African apes. Wood and Collard’s (1999) review of the literature led them to conclude that the developmental schedules of *H. ergaster* and *H. neanderthalensis* were more similar, if not identical, to that of *H. sapiens*, whereas the developmental schedules of *H. habilis* and *H. rudolfensis* were more like those of living chimpanzees and gorillas. Wood and Collard (1999) did not discuss the developmental schedules of *H. erectus* and *H. heidelbergensis* because at the time their paper went to press no comparative analysis of hominid development had included specimens of these species.

It is now clear that Wood and Collard’s (1999) conclusions regarding the life history strategies of the fossil *Homo* species need to be modified. While there are differences between what can be determined about the growth trajectory of Neanderthals and the growth trajectory of modern humans (Thompson and Nelson 2000; Ramirez Rozzi and Bermudez de Castro 2004), the developmental schedule of *H. neanderthalensis* still appears to have been more modern humanlike than apelike (Dean et al. 2001). Likewise, the developmental schedules of *H. habilis* and *H. rudolfensis* still appear to have been more apelike than modern humanlike (Dean et al. 2001). However, the hypothesis that the maturation period of *H. ergaster* was modern humanlike no longer appears tenable. Dean et al.’s (2001) comparative analysis of fossil hominid dental incremental markings suggests that while the pattern of development in *H. ergaster* is similar to the pattern of development in *H. sapiens*, the rate at which *H. ergaster* developed was more apelike than modern humanlike. In addition to altering the assessment of the developmental schedule of *H. ergaster*, work published since Wood and Collard’s (1999) study appeared has shed light on the life history strategy of *H. erectus*. Dean et al. (2001) included a specimen assigned to *H. erectus* in their study. They concluded from this specimen, Sangiran S7-37, that *H. erectus* reached maturity relatively rapidly. *H. erectus* was also found to have had an
apelike pattern of brain growth in a study reported by Coqueugniot et al. (2004) in which the infant *H. erectus* specimen from Java, Perning 1, was compared with a sample of modern humans and chimpanzees. A number of recent studies have examined development in *H. heidelbergensis* (Bermudez de Castro and Rosas 2001; Bermudez de Castro et al. 2003; Ramirez Rozzi and Bermudez de Castro 2004). Unfortunately, these studies have been carried out in such a way that it is difficult to assess with confidence whether the developmental schedule of *H. heidelbergensis* was more like that of *H. sapiens* than those of the great apes or vice versa. Nonetheless, the fact that Ramirez Rozzi and Bermudez de Castro (2004) find enamel extension rates to be slower in *H. heidelbergensis* than in Neanderthals suggests that *H. heidelbergensis* was more modern humanlike than apelike in its developmental schedule. Thus, it now appears that the developmental schedules of *H. heidelbergensis* and *H. neanderthalensis* were more similar to the developmental schedule of *H. sapiens* than to those of chimpanzees and gorillas, whereas the developmental schedules of *H. erectus, H. ergaster, H. habilis, H. rudolfensis* were more like those of chimpanzees and gorillas.

The other area in which there have been developments that affect Wood and Collard’s (1999) conclusions regarding the adaptive strategies of the species conventionally assigned to genus *Homo* is locomotor behavior. Their case for removing *H. habilis* from genus *Homo* was based, in part, on the fact that they considered it to have been a facultative biped like the australopiths rather than an obligate biped like *H. ergaster, H. erectus, H. heidelbergensis, H. neanderthalensis,* and *H. sapiens.* The locomotor behavior of *H. rudolfensis* was not considered because in the absence of an associated skeleton nothing is known for certain about its postcranial morphology. Wood and Collard (1999) cited three lines of evidence in support of their hypothesis. One of these was the morphology of the hand bones associated with OH 7, the type specimen of the species, which have been interpreted as being consistent with an apelike climbing ability (Susman and Creel 1979). Another was the configuration of the semicircular canals in the Southern African *H. habilis* specimen Stw 53. This is so markedly different from the configuration in *H. sapiens* (Spoor et al. 1994) that it is likely the two taxa had different balancing requirements. The third line of evidence Wood and Collard (1999) cited in support of the hypothesis that *H. habilis* was a facultative biped was the limb proportions of the two associated skeletons, OH 62 and KNM-ER 3735, that have been assigned to *H. habilis.* These had been reconstructed as being more primitive than those of *A. afarensis* (Hartwig-Scherer and Martin 1991). Humerus length is similar in modern humans and living chimpanzees, but the former have shorter forearms than the latter. They also have markedly longer femurs. These differences in limb proportions are thought to be related to the contrasting locomotor strategies of the two species: the long femurs of modern
humans being adaptive for bipedalism and the long forearms of living chimpanzees being adaptive for climbing. The available evidence pertaining to limb proportions in australopiths suggests that their forearms were comparatively long while their femora were intermediate in length between those of humans and chimpanzees. This is consistent with the hypothesis that australopiths combined facultative terrestrial bipedalism with proficient climbing. Thus, Hartwig-Scherer and Martin’s (1991) finding that the limb proportions of OH 62 and KNM-ER 3735 were more primitive than those of *A. afarensis* reinforced the hypothesis that *H. habilis* was a facultative biped.

Since Wood and Collard’s (1999) study went to press, a number of analyses have been published that challenge the hypothesis that the limb proportions of *H. habilis* were more primitive than those of the australopiths. For example, in 2002, Brian Richmond, Leslie Aiello, and Bernard Wood reported the results of a randomization-based study designed to examine the significance of the limb proportion differences among several early hominids, including OH 62 and the only *A. afarensis* associated skeleton, AL 288-1. They found that the limb proportions of OH 62 are not statistically significantly different from those of AL 288-1. Thus their analyses did not support the hypothesis. More recently, Reno et al. (2005) have argued that the humerofemoral index of OH 62 cannot be calculated because the portion of femur it retains—the proximal part—is a poor predictor of maximum femur length. This claim is supported by an analysis of the relationship between proximal and maximum femur length in extant hominoids, which suggests that the two lengths are not significantly correlated. Although some researchers have taken issue with the methods and assumptions used by Reno et al. (2005), the lack of association between proximal and maximum femur length in their sample of humans and apes certainly suggests that current estimates of the length of OH 62’s femur or of its humerofemoral index must be treated with caution. Thus, there is no longer any support for the claim that the limb proportions of OH 62 are more primitive than those of the australopiths; the most that can be said is that they are australopith-like.

The developments that have taken place over the last few years in relation to fossil hominid life histories and locomotor abilities have certainly challenged Wood and Collard’s (1999) conclusions regarding the adaptive strategies of some of the species conventionally assigned to genus *Homo*. Most notably, they suggest that *H. erectus* and *H. ergaster* were less modern humanlike than Wood and Collard’s (1999) analyses suggested. However, on balance, the available evidence still suggests that the adaptive strategies of *H. habilis* and *H. rudolfensis* were different from those operated by *H. erectus*, *H. ergaster*, *H. heidelbergensis*, *H. neanderthalensis*, and *H. sapiens* (Table 8.5). Taken together with the results of the phylogenetic studies that have been published in the last 7 years, this
suggests that Wood and Collard’s (1999) proposal to remove *H. habilis* and *H. rudolfensis* from *Homo* and assign them to a different genus or pair of genera remains valid.

The last issue we will consider here is the inclusion of the collection of Late Pleistocene specimens from the site of Liang Bua, Flores, in genus *Homo* as a new species, *H. floresiensis* (Brown et al. 2004). In justifying this course of action, Brown et al. (2004) cite Wood and Collard’s (1999) study in such a way that it appears that incorporating the *H. floresiensis* hypodigm in *Homo* is compatible with the latter authors’ proposals regarding the definition and composition of the genus. However, it is not clear that this is in fact the case. Currently, it is not possible to determine whether or not *H. floresiensis* fulfils Wood and Collard’s (1999) first criterion for including a species within the genus *Homo*, namely that it is more closely related to *H. sapiens* than to the type species of any other genus. This is because the phylogenetic relationships of *H. floresiensis* have yet to be formally analyzed. Some evidence relevant to Wood and Collard’s (1999) second criterion—that the adaptive strategies of fossil species assigned to *Homo* should be more similar to that of *H. sapiens* than to the adaptive strategies of the type species of other genera—is available for *H. floresiensis*, but this evidence is difficult to interpret.

The individual represented by the partial associated skeleton LB1 has been estimated to have been around 106 cm in height and to have weighed between 16 and 36 kg, depending on the body mass proxy employed (Brown et al. 2004). A tibia belonging to another individual (LB8) is consistent with a stature of 109 cm (Morwood et al. 2005). These figures suggest that *H. floresiensis* was more similar in size to the australopiths than it was to *H. sapiens*. It also appears that *H. floresiensis* had a relatively small brain. When Wood and Collard’s (1999) approach to computing relative brain size (cube root of brain size divided by square root of orbital area, product multiplied by 10) is employed, *H. floresiensis* has a relative brain size that is smaller than those of *A. africanus*, *P. aethiopicus*, and *P. boisei* (Table 8.2). Initially, the postcranial anatomy of *H. floresiensis* was suggested to be consistent with the type of obligate bipedalism seen in modern humans as opposed to the form of facultative bipedalism that most researchers believe the australopiths employed (Brown et al. 2004). However, new specimens from Liang Bua, including additional elements of the LB1 associated skeleton, have cast on this hypothesis (Morwood et al. 2005). It now appears that the humerofemoral index of *H. floresiensis* is more similar to that of *A. afarensis* than it is to that of *H. sapiens* (Morwood et al. 2005). Likewise, Morwood et al. (2005) have suggested that the ilium of LB1 is consistent with a teardrop-shaped thorax as seen in the australopiths rather than the barrel-shaped thoracic region
Morwood et al. (2005) also report that LB1’s femoral robusticity falls in the chimpanzee range, and that its humeral robusticity is midway between the chimpanzee range and the human range. These observations suggest that the locomotor behavior of *H. floresiensis* may have been more like that of the australopiths than that of modern humans. However, LB1 apparently differs from all other known hominid species in humeral torsion and aspects of ulna morphology (Morwood et al. 2005), which suggests that this hypothesis may also need to be revised in future. Thus, in terms of body size, relative brain size, and inferred locomotor behavior, *H. floresiensis* appears to be more similar to the australopiths than to the species that Wood and Collard (1999) assign to genus *Homo*.

The available data pertaining to masticatory morphology presents a different picture. Wood and Collard (1999) assessed masticatory system similarities among the hominids on the basis of Euclidean distances derived from 11 size-corrected dental and mandibular variables. So far, data for only six of these variables have been published for *H. floresiensis* (Brown et al. 2004). When Wood and Collard’s analysis is replicated with the six variables, the Euclidean distance between *H. floresiensis* and *H. sapiens* is 1.77, while the comparable distances between *H. floresiensis* and *A. africanus* and between *H. floresiensis* and *P. robustus* are 4.97 and 5.72, respectively (Tables 8.3 and 8.4). Thus, the masticatory apparatus of *H. floresiensis* is more similar in size to that of *H. sapiens* than it is to the type

<table>
<thead>
<tr>
<th>Table 8.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative brain size</td>
</tr>
<tr>
<td><strong>Taxon</strong></td>
</tr>
<tr>
<td><em>P. aethiopicus</em></td>
</tr>
<tr>
<td><em>P. boisei</em></td>
</tr>
<tr>
<td><em>A. africanus</em></td>
</tr>
<tr>
<td><em>H. habilis</em></td>
</tr>
<tr>
<td><em>H. ergaster</em></td>
</tr>
<tr>
<td><em>H. rudolfensis</em></td>
</tr>
<tr>
<td><em>H. heidelbergensis</em></td>
</tr>
<tr>
<td><em>H. erectus</em></td>
</tr>
<tr>
<td><em>H. neanderthalensis</em></td>
</tr>
<tr>
<td><em>H. sapiens</em></td>
</tr>
<tr>
<td><em>H. floresiensis</em></td>
</tr>
</tbody>
</table>

Values taken from Wood and Collard (1999) apart from those for *H. floresiensis*. The brain size figure for *H. floresiensis* was taken from Falk et al. (2005). The orbital area figure for this species was obtained by multiplying the values for orbital height and width provided by Brown et al. (2004). Relative brain size was computed by dividing the cube root of absolute brain size by the square root of orbital area and multiplying the product by 10, as per Wood and Collard (1999).
while the fact that the phylogenetic relationships of \textit{H. floresiensis} have yet to be formally assessed precludes a satisfactory assessment of its attribution to \textit{Homo} as per Wood and Collard’s (1999) definition of the genus, the available data on its adaptive strategy clearly suggest that there is a problem. Given that some of its adaptive characteristics are consistent with those seen in the other species assigned to genus \textit{Homo} by Wood and Collard (1999) while others are not (\textit{Table 8.5}), either Wood and Collard’s (1999) approach to defining \textit{Homo} needs to be amended or \textit{H. floresiensis} needs to be assigned to a different genus. In

\begin{table}
\centering
\caption{Masticatory system relative size}
\begin{tabular}{lcccccccc}
\hline
Taxon & 1 & 2 & 3 & 4 & 5 & 6 & OA \\
\hline
\textit{A. africanus} & 41 & 20 & 33 & 23 & 12.9 & 14.1 & 838.95 \\
\textit{P. boisei} & 51 & 29 & 42 & 29 & 15.7 & 18.5 & 1114.26 \\
\textit{P. robustus} & 50 & 28 & 39 & 27 & 14.1 & 15.7 & 1066.42 \\
\textit{H. erectus} & 37 & 19 & 36 & 22 & 12.0 & 12.7 & 1225.33 \\
\textit{H. ergaster} & 33 & 20 & 31 & 19 & 11.6 & 12.3 & 1180.20 \\
\textit{H. habilis} & 27 & 19 & 29 & 21 & 12.3 & 12.6 & 907.68 \\
\textit{H. neanderthalensis} & 42 & 15 & 34 & 18 & 10.7 & 10.7 & 1403.98 \\
\textit{H. rudolfensis} & 36 & 23 & 36 & 23 & 13.2 & 13.7 & 1084.16 \\
\textit{H. sapiens} & 34 & 14 & 29 & 13 & 10.5 & 10.5 & 1289.37 \\
\textit{H. floresiensis} & 28 & 15 & 20.5 & 15.5 & 11.4 & 10.0 & 992.00 \\
\hline
\end{tabular}
\end{table}

Values taken from Wood and Collard (1999) apart from those for \textit{H. floresiensis}. Figures for the mandibular and dental characters for \textit{H. floresiensis} were obtained from Brown et al. (2004). The orbital area value for this species was obtained by multiplying the values for orbital height and width provided by Brown et al. (2004). (1) Symphyseal height, (2) symphyseal breadth, (3) corpus height at M1, (4) corpus width at M1, (5) M1 buccolinguall diameter, (6) M2 buccolinguall diameter, (OA) orbital area.

\begin{table}
\centering
\caption{Normalized Euclidean distances between fossil \textit{Homo} species and \textit{H. sapiens}, \textit{A. africanus}, and \textit{P. robustus} based on the masticatory system variable values}
\begin{tabular}{lcccc}
\hline
 & \textit{H. sapiens} & \textit{A. africanus} & \textit{P. robustus} \\
\hline
\textit{H. rudolfensis} & 3.96 & 1.75 & \textbf{1.17} \\
\textit{H. habilis} & 3.45 & \textbf{2.63} & 3.40 \\
\textit{H. erectus} & \textbf{2.81} & 2.91 & 3.59 \\
\textit{H. ergaster} & 1.98 & 3.57 & 4.22 \\
\textit{H. neanderthalensis} & 1.19 & 4.54 & 5.19 \\
\textit{H. floresiensis} & 1.77 & 4.97 & 5.72 \\
\hline
\end{tabular}
\end{table}

The figure in bold in each row is the shortest distance.

species of the other two hominid genera for which Wood and Collard (1999) provide data.
our view, the latter course of action is preferable until the phylogenetic relationships and adaptive strategy of the species have been more fully evaluated.

### 8.5 Concluding remarks

Some will undoubtedly view the foregoing as wasted effort. There is a widespread belief that hominid systematics is arcane and irrelevant. However, this latter notion is ill founded. Sound taxonomic units are a prerequisite for progress in evolutionary biology (Crowson 1970; Panchen 1992). Thus more attention should be paid to the systematics of the hominids, not less. With this in mind, we hope that the points we have made in this chapter stimulate further work on the definition and composition of the genus *Homo*. In particular, there is a pressing need for Mayr's (1950) and Cela-Conde and Altaba's (2002) approaches to defining genera to be operationalized satisfactorily. We also badly need reliable information about both the phylogenetic relationships of the early *Homo* species and their postcranial morphology, especially as it relates to locomotion. Lastly, it would be helpful for the systematic comparative approach to analyzing dental development employed by Dean et al. (2001) to be extended to *H. heidelbergensis* and to the fossils assigned to the taxon that, for the time being at least, is referred to as *H. floresiensis*.

### Acknowledgments

We thank Nicole Collard, Alan Cross, Stephen Lycett, and Daniele Serdoz for providing helpful comments on earlier drafts of this chapter.
References


Busk G (1865) On a very ancient human cranium from Gibraltar. Report of the 34th Meeting of the British Association for the Advancement of Science (Bath 1864), pp 91–92


Dean MC, Leakey MG, Reid D, Schrenk F, Schwartz GT, Stringer CB, Walker A (2001)


Flourens P (1856) *Cours de Physiologie Comparée: de l’Ontologie ou étude des êtres*. Baillière, Paris


Groves CP (2001a) Primate taxonomy. Smithsonian Institution Press, Washington DC


Schoetensack O (1908) Der Unterkiefer des Homo heidelbergensis aus den Sanden von Mauer bei Heidelberg. W Engelmann, Leipzig
Defining the genus Homo

9 The Earliest Putative Homo Fossils

Friedemann Schrenk · Ottmar Kullmer · Timothy Bromage

Abstract

The earliest fossil remains of the genus Homo have been discovered in eastern, southeastern, and southern Africa. The sample comprises about 200 skeletal fragments attributable to about 40 individuals and assigned to two species: Homo rudolfensis (2.5–1.8 Ma) showing a combination of primitive dentition with Homo-like locomotion and Homo habilis (2.1–1.5 Ma) exhibiting a progressive reduction of tooth roots but resembling great apes rather than humans in the postcranial skeleton. Another significant difference between early Homo and the australopithecines is the brain size, which was larger in early Homo than Australopithecus but smaller than in Homo erectus. Endocasts of H. habilis from Olduvai Gorge and Koobi Fora reveal a number of distinctive features some of which are recognized as Homo autapomorphies. Differences in tooth wear between H. rudolfensis, with megadont teeth and more horizontal tooth abrasion, and H. habilis, with more gracile molars and higher relief in worn teeth indicate significant differences in diet and ecology of early Homo species. The origin of the genus Homo coincided with the onset of material culture. Between ca. 2.8 and 2.5 Ma, extensive open habitats comprising more arid-tolerant vegetation developed in Africa. The selective pressures of this habitat change resulted in the increased survival of more megadont species varieties. Megadonty allowed these species to feed on tougher fruit and open woodland–open savannah food items resulting in the phyletic splitting of Australopithecus afarensis into Paranthropus and Homo lineages by ca. 2.5 Ma. An evolutionary scenario that complies with both the Habitat Theory and early hominid biogeography is provided. It delineates the association between faunal turnover and climate change, and suggests a single origin for the Paranthropus lineage but separate origins for H. rudolfensis and H. habilis from A. afarensis and A. africanus ancestors, respectively.
9.1 Introduction

The search for the roots of the genus *Homo* is of particular interest in the field of paleoanthropology. The taxonomic determination of the earliest putative *Homo* fossils provides the basis for the definition of the taxon *Homo* to which all modern humans belong.

Carolus Linnaeus (1758) established the genus *Homo* in the tenth revision of his *Systema Naturae*. In his opinion, *Homo* subsumed six groups: *H. sylvestris*—a mixture of orangutan and myths, *H. sapiens*, and four geographical variants from Africa, America, Asia, and Europe. In the two centuries that followed *H. neanderthalensis* (King 1864), *H. heidelbergensis* (Schoetensack 1908), *H. erectus* (Dubois 1892; Mayr 1944), *H. habilis* (Leakey et al. 1964), *H. ergaster* (Groves and Maza´k 1975), *H. rudolfensis* (Alexeev 1986), *H. antecessor* (Bermudez de Castro et al. 1997), and others were included in the genus *Homo*.

The history of research, the order of discoveries, and existing paradigms heavily influence the formation and change in interpretations of human evolution. This holds true especially regarding ideas on the origin of the genus *Homo* since the 1960s. Debates on the attribution of fossil specimens and the definition of the genus *Homo* continue today (Wood 1992).

Opinions differ regarding the number of species and also the specimens included in the genus. Some even assign all putative *Homo* specimens to living humans (*H. sapiens*). Species names in paleoanthropology are labels rather than natural species, and the taxonomic determination of fossils is more or less a question of the philosophy followed by the authors. There are both theoretical and practical reasons to erect taxa, as chronospecies for time equivalent appearance or morphospecies for a complex of shared anatomical features, and in the worst case there are political reasons for the allocation of species.

9.2 Fossil evidence

Early research on the origin of the genus *Homo* is closely related to the African fieldwork of Louis Leakey (1903–1972). He strongly believed in Africa as the cradle of humankind and in 1932 discovered the first evidence of early *Homo* at Kanam, (Kenya) east of Lake Victoria—a specimen, which today is attributed to *H. erectus*. He also undertook archeological surveys in Olduvai Gorge, Tanzania (Figure 9.1), where later he discovered early pebble tools in Bed I (ca. 1.8 Ma)—remains of what he termed the “Oldowan” industry. The search for the artifacts creator led to the discovery of robust australopithecine remains(*Zinjanthropus boisei*) (Leakey 1959). However, due to its small brain size,
A year later, Jonathan Leakey, the son of Louis, discovered two fragments of a relatively gracile skull, a lower jaw (Figure 9.2), and several hand bones of OH 7 (Leakey 1961), deriving from the same stratigraphic level (Bed 1) in Olduvai. Brain volume was estimated at around 680 cm$^3$, a significantly higher value than in robust australopithecines. Consequently, this fossil was interpreted as representing a progressive hominid type of unknown species affiliation. Later Leakey et al. (1964) decided on a new species *H. habilis*. Raymond A. Dart, the founder of modern paleoanthropology in Africa, who in 1925 had introduced the genus *Australopithecus*, had suggested this name to them. The Latin term “habilis” means
“handy, skillful, able”: finally, the producer of the Oldowan culture seemed to have been identified.

Apart from OH 7, the species description of *H. habilis* included skull fragments and teeth (OH 4 and OH 6), part of an adult foot (OH 8), and the incomplete skull of an adolescent (OH 13). Further, Leakey et al. (1964) referred a collection of juvenile cranial pieces (OH 14) and the fragmented cranial vault and dentition (OH 16) of a young adult to the new species.

Since then, numerous additional fossils of *H. habilis* have been discovered at Olduvai Gorge: 9 fragmentary skulls, 4 mandible fragments, 19 teeth, and 8 postcranial fragments. Among these fragments was the squashed skull OH 24 (Twiggy), which was found in 1968. In 1986, a partial female skeleton (OH 62) was assembled from a number of fragments (Johanson et al. 1987). This specimen showed that *H. habilis* was fully bipedal and had a brain larger than all australopithecines. For many years, the *H. habilis* remains from Olduvai Gorge were seen as the most important early *Homo* specimens and consequently played the leading role in most hypotheses regarding the origin of genus *Homo*.
Leakey et al. (1964) originally discussed cranial and mandibular traits, to distinguish the *Homo* specimens of Olduvai from Australopithecines and *H. erectus*. Maxillary and mandibular size is smaller than in *Australopithecus* and tends in size to *H. erectus* and *H. sapiens*. The surface of the skull shows slight to strong muscular markings and the parietal curvature in the sagittal plane varies from slight to moderate. The frontal bone is more vertical and the torus supraorbitalis is less developed than in australopithecines. In the occipital region, the relatively open-angled external sagittal curvature differs markedly from *Australopithecus*.

In 1970 the picture of earliest *Homo* began to change significantly with the success of the Koobi Fora Research Project in northern Kenya and led by Richard Leakey, Louis’ son. In just a few years on the eastern shores of Lake Turkana, his team recovered 9 skulls, 10 mandibles, 6 isolated teeth, and 5 postcranial fragments (Leakey 1973a,b). Originally, all of the early *Homo* finds from East Turkana, with an age similar to those from Olduvai (1.9–1.8 Ma), were interpreted as bearing similarities to *H. habilis*—then the only early species of *Homo*. One cranial fragment from the Nachukui Formation on the western shores of Lake Turkana was also assigned to *H. habilis*. However, two of the best preserved skulls from Koobi Fora (KNM-ER1470, Figure 9.4; KNM-ER 1813, Figure 9.5) later gave rise to an extended debate among researchers about the heterogeneity of the *H. habilis* hypodigm, and finally led to the recognition of a new species, *H. rudolfensis*.

In the 1970s a large number of isolated *Homo* teeth were discovered north of Koobi Fora, near the Omo-River in southern Ethiopia, in Member G and H of the Shungura Formation. From these it became clear that the origin of the genus *Homo* extended well beyond 2 Ma. Already in 1965, a temporal bone had been discovered by John Martyn at Chemero, Kenya, which nearly two decades later was described by Hill et al. (1992) as a very early member of the genus *Homo*, dated to around 2.4 Ma.

In 1976, an early *Homo* fossil was found at Sterkfontein, South Africa (Stw 53), which belonged neither to *H. erectus* nor to *Australopithecus* (Hughes and Tobias 1977). A partial facial skull (SK 847 from Swartkrans), assembled from several fragments, originally attributed to a different species, is further evidence for *H. habilis*, which probably migrated into southern Africa around 2 Ma (see below).

The geographical gap between the southern and eastern African early hominid sites was filled in the early 1990s through discoveries in the “Hominid Corridor” of the northern Malawi Rift (Schrenk et al. 1993; Bromage and Schrenk 1995). In 1992, at Uraha, the Hominid Corridor Research Project (HCRP) has
recovered a mandibular corpus, UR 501 (Figure 9.3), containing third and fourth premolars and first and second molars in variable states of preservation (Schrenk et al. 1993). Many absolute and relative measures defining molar and premolar crown shape indices, relative cusp areas, fissure patterns and enamel microanatomical features, as well as overall crown morphology, are within the sample range of early Homo, although some may also be subsumed within the limits of variation represented by Australopithecus (A. africanus and A. afarensis). However, UR 501 has absolutely large molar crown areas, relative expansion of the P3 talonid, plate like P3 and P4 roots, and some enamel microanatomical features correspond more closely to the Paranthropus condition. UR 501 corresponds closely to the subset of Late Pliocene fossils from east Turkana, Kenya, which demonstrates relatively large brains and robust jaws and teeth and based on the above have been assigned to H. rudolfensis by Alexeev (1986) and Wood (1992), and to which UR 501 was also referred (Bromage et al. 1995). This Malawi specimen has been dated by biostratigraphic correlation of suid material with well-dated sites in southern Ethiopia (Omo Shungura) and northern Kenya (Koobi Fora), indicating an age of about 2.4–2.5 Ma whereas most early Homo fossils are around 2 Myr old.

Figure 9.3
UR 501 from the Chiwondo Beds, northern Malawi (ca. 2.5–2.4 Ma), Homo rudolfensis (Drawing: Claudia Schnubel)
It is important to note that the appearance of earliest Homo is contemporaneous with the origin of hyperrobust australopithecines (Paranthropus). The earliest evidence for this co-existence is based on further hominid discoveries in the Chiwondo Beds of Northern Malawi. A maxillary fragment (RC 911) preserves part of the left alveolar process, with badly damaged M¹ crown and fragmentary M² crown. Size, morphology and abrasion occlusal pattern on the surface suggest that RC 911 should be assigned to Paranthropus boisei (Kullmer et al. 1999).

The biogeographic significance of these Malawi Rift hominids lay in their association with the eastern African endemic faunal group. The associated bovid and suid faunas show a small amount of overlap with southern African animals and a greater overlap with eastern African faunal elements. Biogeographic variation, in the Malawi Rift may be linked to habitat change occurring due to climate shifts, with maximum change occurring around 2.5 Ma.

9.3 Changing taxonomy

Hominid fossils are generally assigned to the genus Homo if they fulfill four main criteria (Keith 1948; Tobias 1991; Wood and Collard 2001): a brain size above 600 cm³, ability for speech and tool-making, as well as an opposable pollux. To date, the hypodigm of earliest Homo attributed to H. habilis sensu stricto and H. rudolfensis contains about 200 skeletal fragments attributable to about 40 individuals (Tables 9.1 and 9.2). Despite—or even due to—the large number of specimens, the taxonomic interpretation of earliest Homo is highly controversial (Wood 2000).

Originally, the interpretation of the early Homo hand as “modern” (Leakey et al. 1964) supported the view of H. habilis as an early but “able” human as opposed to the rather “clumsy” australopithecines. However, later skeletal finds at Olduvai Gorge (OH 62) (Johanson et al. 1987) demonstrated that the postcranial skeleton of H. habilis indeed resembled Australopithecus africanus rather than Homo. Yet the most distinctive character of H. habilis remains its relatively and absolutely higher brain volume compared to that of Australopithecus. The forehead of H. habilis is more vertical and a weak supraorbital torus is present. Whereas the morphological characters are quite uniform in the Olduvai sample, the discussion started to heat up mainly over two very distinct skull fragments from Koobi Fora: KNM-ER 1470 (Figure 9.4) (Leakey 1973a) and KNM-ER 1813 (Figure 9.5) (Leakey 1973b).

In a comprehensive character analysis of all available putative H. habilis fossils from Koobi Fora, Wood (1991) concluded that the variability exhibited
by the sample was not only the result of sexual dimorphism as was suspected at the time, but that highly significant differences exist throughout the entire skeleton. There is a mosaic of *Australopithecus* and *Homo* characters in both early species: Whereas *H. rudolfensis* exhibits a combination of ancestral dentition with *Homo*-like locomotion, *H. habilis* shows a progressive reduction of tooth roots and resembles great apes rather than humans postcranially.

### Table 9.1

#### Significant morphological differences between *H. habilis* and *H. rudolfensis*

<table>
<thead>
<tr>
<th></th>
<th><em>Homo habilis sensu stricto</em></th>
<th><em>Homo rudolfensis</em></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Skull and teeth</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Absolute brain size (cm³)</td>
<td>An average volume of 610</td>
<td>An average volume of 751</td>
</tr>
<tr>
<td>Overall cranial vault morphology</td>
<td>Enlarged occipital contribution to the sagittal arc</td>
<td>Primitive condition</td>
</tr>
<tr>
<td>Endocranial morphology</td>
<td>Primitive sulcal pattern</td>
<td>Frontal lobe asymmetry</td>
</tr>
<tr>
<td>Suture pattern</td>
<td>Complex</td>
<td>Simple</td>
</tr>
<tr>
<td>Frontal</td>
<td>Incipient supraorbital torus</td>
<td>Torus absent</td>
</tr>
<tr>
<td>Parietal</td>
<td>Coronal &gt; sagittal chord</td>
<td>Primitive condition</td>
</tr>
<tr>
<td>Face-overall</td>
<td>Upper face &gt; midface breadth</td>
<td>Midface &gt; upperface breadth: markedly orthognathic</td>
</tr>
<tr>
<td>Nose</td>
<td>Margins sharp and everted; evident nasal sill</td>
<td>Less everted margins; no nasal sill</td>
</tr>
<tr>
<td>Malar surface</td>
<td>Vertical or near vertical</td>
<td>Anteriorly inclined</td>
</tr>
<tr>
<td>Palate</td>
<td>Foreshortened</td>
<td>Large</td>
</tr>
<tr>
<td>Upper teeth</td>
<td>Probably two-rooted premolars</td>
<td>Premolars three-rooted; absolutely and relatively large anterior teeth</td>
</tr>
<tr>
<td>Mandibular fossa</td>
<td>Relatively deep</td>
<td>Shallow</td>
</tr>
<tr>
<td>Foramen magnum</td>
<td>Orientation variable</td>
<td>Anteriorly inclined</td>
</tr>
<tr>
<td>Mandibular corpus</td>
<td>Moderate relief on external surface; rounded base</td>
<td>Marked relief on external surface; everted base</td>
</tr>
<tr>
<td>Lower teeth</td>
<td>Buccolingually narrowed; postcanine crowns; reduced talonid on P₄; M₃ reduction; mostly single-rooted mandibular premolars</td>
<td>Broad postcanine crowns; relatively large P₄ talonid; no M₃ reduction; twin, platelike P₄ roots, and bifid, or even twin, platelike P₃ roots</td>
</tr>
<tr>
<td>Postcranium</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limb proportions</td>
<td>Apelike</td>
<td>?</td>
</tr>
<tr>
<td>Forelimb robusticity</td>
<td>Apelike</td>
<td>?</td>
</tr>
<tr>
<td>Hand</td>
<td>Mosaic of apelike and modern humanlike features</td>
<td>?</td>
</tr>
<tr>
<td>Hindfoot</td>
<td>Retains climbing adaptations</td>
<td>Later <em>Homo</em>-like</td>
</tr>
<tr>
<td>Femur</td>
<td>Australopithecine-like</td>
<td>Later <em>Homo</em>-like</td>
</tr>
</tbody>
</table>

Based on the work of Wood (1991) and subsequent re-evaluation of fossils it is evident that two distinct types can be separated. Whereas one group, represented by KNM-ER 1813, follows the original description of *H. habilis* from Olduvai Gorge, a new group is represented by KNM-ER 1470, for which no comparison existed in Olduvai at the time. Although the subsequent find of OH 65 at Olduvai (Blumenschine et al. 2003) showed a mixture of *H. habilis sensu stricto* and KNM-ER 1470 features, it is still a valid conclusion that about half of the early *Homo* material from Koobi Fora belongs to *H. habilis sensu stricto*, with an age of 2.1–1.5 Ma, which includes also specimens from Koobi Fora, West-Turkana, Omo, Olduvai Gorge, and southern Africa (Table 9.2). Hominid remains from Ubeidiyah in Israel (Leakey et al. 1964) and *Meganthropus palaeojavanicus* from Java, Indonesia, which at one stage were tentatively assigned to *H. habilis* (Tobias and von Koenigswald 1964), are not considered as such today.

The Koobi Fora early *Homo* material not assigned to *H. habilis* is defined as the more recently named species *H. rudolfensis*, with an age of 2.5–1.8 Ma, which also includes specimens from Chemeron, West-Turkana (Prat et al. 2005), and Omo (Suwa et al. 1996; Ramirez Rozzi 1997) as well as northern Malawi (Schrenk et al. 1993) (Table 9.3).

| Table 9.2 |
| Fossil remains of *Homo habilis sensu stricto* |

*Homo habilis* (better-preserved specimens in bold)

<table>
<thead>
<tr>
<th>Sites</th>
<th>Skulls and crania</th>
<th>Mandibles</th>
<th>Isolated teeth</th>
<th>Postcranial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Olduvai (OH)</td>
<td>6, 7, 13, 14, 16, 24, 52, 62</td>
<td>7, 13, 37, 62</td>
<td>4, 6, 15, 16, 17, 21, 27, 31, 32, 39, 40, 41, 42, 44, 45, 46, 47, 55, 56</td>
<td>7, 8, 10, 35, 43, 48, 49, 50, 62</td>
</tr>
<tr>
<td>Koobi Fora (KNM-ER)</td>
<td>807, 1478, 1805, 1813, 3735</td>
<td>1501, 1502, 1506, 1805, 3734</td>
<td>808, 809, 1462, 1480, 1508, 1814</td>
<td>813, 1472, 1481, 3228, 3735</td>
</tr>
<tr>
<td>Omo</td>
<td>L894-1</td>
<td>Omo 222-2744</td>
<td>L28-31; L398-573, 1699; Omo 33-3282, Omo 47-47; Omo 74-18; Omo 123-5495; Omo 166-781; Omo K7-19; Omo SH1-17; P933-1</td>
<td></td>
</tr>
<tr>
<td>West Turkana</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Sterkfontein</td>
<td>Stw 53, SE 255, 1508, 1579, 1937, 2396; Sts 19</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Swartkrans</td>
<td>SK 847 (?)</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>
The species name *rudolfensis* was coined by Russian paleontologist V.A. Alexeev, who in 1986 described KNM-ER 1470 as “*Pithecanthropus rudolfensis*,” after Lake Rudolf, the name of Lake Turkana prior to Kenyan independence in 1963.

### 9.4 Postcranial skeleton

The postcranial skeleton of *H. habilis* was characterized by Leakey et al. (1964) using a number of characteristics: the clavicle resembles that of *H. sapiens*, the hand shows broad terminal phalanges, and capitate and MCP articulations also resemble *H. sapiens*, but differ in respect of the scaphoid and trapezium, attachments of the superficial flexor tendons, and the robusticity and curvature of the phalanges. The foot bones resemble *H. sapiens* in the stout and adducted hallux and well-defined foot arches, but differ in shape of the talan trochlea surface and the relatively robust third metatarsal.
All these features indicate an affinity towards *H. sapiens* and underline that the early postcranial material from Olduvai seems to be different from that of the australopithecines. Functional implications based on more detailed descriptions of the foot (OH 8, OH 10) and leg (OH 35), both of which are probably from the same individual (Stern and Susman 1983), were generally more cautious (Wood 1992). According to Stern and Susman (1983), *H. habilis* had not reached the characteristic bipedal gait of *H. sapiens*, since the functional morphology of the knee joint was not well adapted for striding. Further analysis of the OH 8 foot demonstrated that some features, common in nonhuman hominoids, are also present in the foot bones (Lewis 1989). As more postcranial material was uncovered at Koobi Fora, interpretation of hind limb function of early *Homo* became more complex. However, some specimens, such as the femur KNM-ER-1472A and the talus KNM-ER-813, may belong to *H. erectus* or *H. ergaster* rather than to *H. habilis* (Tobias 1991). Although the partial skeleton OH 62 from Olduvai
Gorge was interpreted as evidence for fully upright human bipedal locomotion, Haeusler and McHenry (2004) demonstrated that there is little evidence in support of ancestral body proportions with short legs and long arms in *H. habilis*. Their results suggest that it is more likely that earliest *Homo* possessed an elongation of the legs relative to *A. africanus* and *A. afarensis*, whereas long forearms were still retained (Johanson et al. 1987). With its upper-to-lower limb shaft length proportions, OH 62 falls within the upper range of modern humans and the lower range of chimpanzees due to the partial overlap between these taxa at small body sizes. KNM-ER 3735, the larger-bodied early *Homo* from Koobi Fora, falls well outside chimpanzees and reflects the average proportions of modern humans. Comparison of the Koobi Fora and Hadar postcranial remains leads to the interpretation that *H. habilis* did probably possess a modern pattern of limb shaft proportions and the body proportions of OH 62 are in agreement with other available evidence of *H. habilis* postcranial material (Johanson et al. 1987). The change in the proportions of limb length toward the development of long legs may be indicative of long-distance terrestrial walking in early *Homo* and probably implies a shift in hominid ecology.

### Table 9.3

#### Fossil remains of *Homo rudolfensis*

<table>
<thead>
<tr>
<th>Sites</th>
<th>Skulls and crania</th>
<th>Mandibles</th>
<th>Isolated Teeth</th>
<th>Postcranial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Koobi Fora</td>
<td>1470, 1590, 3732, 3891</td>
<td>819, 1482, 1483, 1801, 1802</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Cheameron</td>
<td>KNM-BC 1</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Omo</td>
<td>Omo 75-14</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

The earliest putative *Homo* fossils

### 9.5 Brain and language

A significant difference between early *Homo* and Australopithecines exists in brain size, which was larger in early *Homo* than in *Australopithecus* but smaller
than *H. erectus*. Endocasts of *H. habilis* from Olduvai and Koobi Fora reveal a number of distinctive features, some of which are recognized as autapomorphies of the genus *Homo*. Tobias (1987) defined the principal morphological trait to distinguish *H. habilis* from *Australopithecus* as a larger mean endocranial capacity in the former (640 cm$^3$) than in *A. africanus* (441 cm$^3$), *A. boisei* (513 cm$^3$), and *A. robustus* (530 cm$^3$). This suggests that the evolutionary trend toward brain expansion was already well under way more than 2 Ma. The *H. habilis* mean (640 cm$^3$) is close to the lower limit (647 cm$^3$) of the 95% population range of *H. erectus* but well above the upper limit of the *A. africanus* range (492 cm$^3$). The brain capacity of the *H. rudolfensis* type specimen (750 cm$^3$) is larger than the known range for *H. habilis* from Olduvai Gorge and Koobi Fora and falls within the lower range of *H. erectus*.

A prominent feature of the *H. habilis* brain is the bilateral transverse expansion of the cerebrum, especially in the frontal and parieto-occipital areas, and a posterior heightening. The increased bulk of cerebral frontal and parietal lobes and the sulcal and gyral pattern of the lateral frontal lobe have been interpreted as derived features for the genus *Homo* (Tobias 1987). The *H. habilis* brain showed a well-developed left superior parietal lobule, and a prominent development of the inferior parietal lobule.

The endocast of KNM-ER 1470 shows a left frontal lobe sulcal pattern that is associated with Broca’s area in living people (Falk 1987), a finding that has led to the conclusion that *H. rudolfensis* may have been capable speech. This conclusion is in accordance with Holloway’s observation of a pronounced left-occipital-right-frontal petalia pattern in the KNM-ER-1470 endocast that may indicate functional cortical asymmetry (Holloway 1983). Surprising corroborative evidence has been provided by Toth’s analyses of stone flakes, which indicate that hominids were predominantly right-handed by 2 Ma (Toth 1985). Adjacent areas in the left frontal lobe control the speech organs and the right hand. Tobias (1991) stated that *H. habilis* is the earliest hominid to show prominent enlargement of Broca’s and Wernicke’s areas. If so, the same should be seen in *H. rudolfensis*. *Australopithecus* endocasts show Broca’s area, but not Wernicke’s region, while anthropoid apes display neither of the two. The prominent development of the two speech areas may thus be seen as an important autapomorphy of the genus *Homo* (Tobias 1991). Even if *H. habilis* and *H. rudolfensis* possessed the neurological basis of speech, there is no evidence that either of them used spoken language. The areas of the brain that control spoken communication probably manifested themselves only when brain enlargement occurred and marked encephalization started.
9.6 Material culture and food processing

Around 2.5 Ma, simultaneously with an increase in drier and harder food stuffs due to increasing aridity in eastern Africa (de Menocal 2004), there occurred the first hyperrobust australopithecines (*Paranthropus*) and the genus *Homo* in the fossil record (Bromage et al. 1995). This demonstrates an evolutionary alternative to the massive masticationary system of *Paranthropus*, which was capable of dealing with very tough foods. The evolutionary alternative to megadonty was the manufacturing and use of stone tools. The oldest chopper tools are known from Ethiopia (Hata, Bouri Formation) and from Tanzania, approximately 2.5 Ma (Kaiser et al. 1995). From Gona, east of Hadar in the Afar Triangle, primitive pebble tools are dated to 2.6 Ma (Harris 1986). New discoveries from west of Lake Turkana confirm the existence of an early tool culture around 2.5 Ma (Roche et al. 2003). In the Hadar area of Ethiopia, stone tools were found associated with early *Homo* remains (Kimbel et al. 1996). At many sites the presence of more than one hominid species, occurring in the same horizon as early Oldowan pebbles does not give clear evidence of who were the first tool-makers. However, the distinct specialization of the skull and dental morphology in robust australopithecines and brain expansion in early *Homo* point to the latter as the most likely tool manufacturer.

Implements are widely used by higher primates. Yet during marked habitat shifts, which led to pronounced changes in food resources, it probably was the invention of stone tools, which supported the origin of the genus *Homo* around 2.5 Ma. Increasing independence from the environment led to an increase in the dependence on culture.

If early *Homo* utilized stone tools to prepare food, the dentition might actually indicate these behavioral changes in food acquisition. However, the morphology of early *Homo* teeth does not seem to suggest extensive food preparation before ingestion. The incisors are large compared to those of *Australopithecus* and *H. erectus*, and the canines are large relative to the premolar crown surfaces. The premolars are narrower than in *Australopithecus* and fall within the range of *H. erectus*. Molar size overlaps the ranges for *Australopithecus* and *H. erectus*. The cheek teeth of *H. rudolfensis* are enlarged and show affinities to *Paranthropus* molars. All teeth are relatively narrow buccolingually and elongated mesiodistally, especially the mandibular molars and premolars. In *H. habilis* we see well-developed third molars, while in *H. rudolfensis* the third molar forms a smaller crown than the second molar. The occlusal surface of the cheek teeth is not as broad as in australopithecine molars and indicates differences in chewing. Tobias (1987) states that the crown’s cusps relief is still present even when the teeth are in advanced wear and dentine is visible. This means that the attrition of
enamel is less pronounced than in earlier hominids. Differences in tooth wear between *H. rudolfensis*, with megadont teeth and a more horizontal tooth abrasion, and *H. habilis*, with its more gracile molars and higher relief in worn teeth are clearly visible. This indicates significant differences in diet and ecology of early *H.* species. *H. rudolfensis* and the robust australopithecines share some cranial and dental features regarding the morphology of the masticatory apparatus (Wood 1992), which indicates that these hominids were able to cope with tough fruits and plants. Since those features are judged as an adaptation to drier climatic conditions, they also show that *H. rudolfensis* was relatively conservative nutritionally and probably followed a more herbivorous strategy in food acquisition.

9.7 Biogeographic scenario

The scenario presented here is derived mainly from hypotheses about early hominid evolution in the context of environmental change and faunal biogeography. It is thus a biogeographic perspective against which we understand the relevance to studies of early *Homo* systematics in general, and morphology and character transformation more specifically.

The behavioral inclination of the earliest hominids distributed along the margin of the tropical rain forest, was to maintain a connection to and remain near the borders between broad riparian habitats and open woodlands during the ascendancy of more warm and humid times. Over short geological timescales, this was typically a local, nondispersing tendency, but by approximately 4 Ma, several species of *Australopithecus* had successfully dispersed throughout the reaches of the African Rift Valley and into western Africa (Figure 9.7). Over longer time frames, this included dispersal through the riparian “corridor” connecting eastern and southern Africa, permitting population dispersal into southern Africa by 3 Ma. This dispersing population maintained habitat specificities to forested environments (Rayner et al. 1993), though in more environmentally temperate climes and in relative geographical isolation at the extreme distal edge of its distribution. The dispersion along changing latitudinal circumstances covaried with its transformation into, first, a geographic variant and, subsequently, into *A. africanus*, joining ranks with other southern African endemic faunas. Thus, *A. afarensis* was essentially an eastern African endemic and it follows that no typical representatives are likely to be recovered from southern African deposits older than 3.5 Ma. (Figures 9.6 and 9.7).

By approximately 2.8 Ma, the initiation of cooler and drier conditions prevailed upon the African landscape, its vegetation, and its faunas, until
climaxing ca. 2.5 Ma (Bonnefille 1980; Vrba 1985, 1988; Prentice and Denton 1988; de Menocal 2004). During this time, A. afarensis in eastern Africa and A. africanus in southern Africa were each subject to unique paleobiogeographic
consequences of this global aridification, in accordance with the “Habitat Theory” of Vrba (1992).

For *A. afarensis*, then, the changing climate meant vicariance of its habitat and its distribution into more removed ecotonal riparian and closed lake margin environs. During the interim between ca. 2.8 and 2.5 Ma, these changing conditions engendered more extensive open habitats comprising more resistant arid-tolerant vegetation around the remaining relatively lush but narrowed “ribbons” of tree-lined riverine forest. The selective pressures of this habitat change resulted in the increased survival of more megadont varieties capable of feeding on tougher fruit and open woodland–open savannah food items. This was so for early hominid as well as numerous eastern and southern African large terrestrial vertebrate lineages ca. 2.5 Ma (Turner and Wood 1993). These pressures were
likewise sufficient to result in the phyletic splitting of A. afarensis into Paranthropus and Homo lineages by ca. 2.5 Ma (Vrba 1988) (Figures 9.6 and 9.7).

Ensuing cooler and drier conditions favored a tougher savannah vegetation composed of plant species better able to retain their moisture under such conditions. Selection favored more facially robust and large molar-toothed mammals, including early hominids, capable of efficiently processing the tougher, more durable vegetation of the savannah. Our evidence suggests that the tropical equatorial animals, including the hominids, of eastern Africa stayed in the tropical African ecological domain, while during the drying and cooling of global climates ca. 2.5 Ma, the southern and more temperate African faunas followed their northward-drifting vegetation zones. Thus, Homo and Paranthropus may have emerged in tropical Africa as a result of the ca. 2.5-Ma climatic cooling event and remained there.

The beginnings of the Paranthropus lineage maintained a reliance on fruiting resources on the riverine side of its ecotone, particularly during the dry season, but it was equally adept at grinding on the postcanine dentition those food items it required from more open habitats during more hospitable times of the year. The beginnings of the Homo lineage, also ca. 2.5 Ma, and represented by H. rudolfensis, was an endorsement of its recency of common ancestry with A. afarensis, a distinction it shared with Paranthropus (Bromage et al. 1995).

However, while Paranthropus was principally adapted by means of a robust masticatory system to its tough and abrasive diet, H. rudolfensis exhibited an increased behavioral flexibility as its adaptation to climatic circumstances included a larger and more provoking, inquiring, and capable brain. This included a shift to proportionately less abrasive foodstuffs and more omnivorous habits. Material culture ameliorated the effects of climate change to the degree that it enabled H. rudolfensis to take advantage of other resources more efficiently than was ever possible before.

P. boisei and H. rudolfensis remained endemic to tropical latitudes during this time (Bromage et al. 1995). The eastern African tropical faunas, having habitable alternatives, remained within their biogeographic domain rather than brave the relative deterioration and paucity of habitats south of the African Rift Valley.

The faunas of southern Africa were subject to a different set of environmental sequelae during the ca. 2.5-Ma cooling event. Waning of the forests and woodlands in deference to more open arid grasslands invigorated not only evolutionary adaptations to savannah life in tropical eastern Africa but also resulted in the distribution drift northward of faunas tracking the equator ward shift of grassland and woodland biomes into eastern Africa from the south ca. 2.5 Ma (Bromage et al. 1995). The temperate zone ca. 2.5 Ma experienced more seasonal
extremes and many organisms unwittingly maintained their inherited preference for moderate seasonal climes and temperate vegetation physiognomy by moving northward with the shrinking of this biome toward the equator, effectively transgressing the Zambezi Ecozone. Among these migrants was *A. africanus* who, having been adapted to a modest temperate ecology, now found its suitable habitats shifted to the north toward the African Rift Valley. While dispersing toward the eastern African tropical domain, selection for increased behavioral flexibility was related to the habitat diversity of the tropics and the presence of other nonvegetative food resources available in their new region. This emerging taxon, *H. habilis*, rapidly established itself as a categorical omnivore and found that it could buffer itself more resolutely from environmental changes. This enabled it to cross habitat boundaries more easily and also to advantage itself of more resources with its material culture.

By approximately 2 Ma, Africa was rebounding from its relatively cool and dry climate to return to slightly more warm and humid conditions (Bromage et al. 1995). A phase of biome expansion ensued that facilitated dispersions away from the equator, ending nearly 1 Myr of relative endemism dominated by tropical equatorial speciations. *P. boisei* dispersed southward along reestablished ecotonal habitats into southern Africa, varied there as a geographic variant under more temperate conditions, and evolved into *Paranthropus robustus* (Figure 9.6). *H. habilis* expanded southward into the southern African temperate domain, but it maintained a very much broader niche and increased its distributional area as a single species. *H. rudolfensis* remained endemic to the eastern African tropical domain due partly to its preference for more open habitats around the rain shadows of the African Rift Valley and partly, perhaps, to some small measure of competitive exclusion from geographic realms occupied by *H. habilis*.

**References**

The earliest putative Homo fossils

and faunal evolution during the Pliocene-
 Pleistocene, Earth Planet Sci Lett 220: 3–24
Dubois E (1892) Palaeontologische anderzoek-
ingen op Java. Verslag van het Mijnwezen.
Batavia 3: 10–14
Rev Anthropol 16: 13–30
Groves CP, Mazák V (1975) An approach to the
taxonomy of the Hominidae: Gracile Villa-
franchian hominids of Africa. Casopis pro
Mineralogii A Geologii 20: 225–247
Haeusler M, McHenry HM (2004) Body propor-
tions of Homo habilis reviewed. J Hum
Evol 46: 433–465
Harris JWK (1986) Découverte de matériel
archéologique oldowayen dans le rift de l'Afar.
L'Anthropologie 90: 339–357
719–722
search for units, models and synthesis. Can J
Anthropol 3: 215–230
Hughes AR, Tobias PV (1977) A fossil skull
probably of the genus Homo from Sterkfon-
Johanson DC, Masao FT, Eck GG, White TD,
Walter RC, Kimbel WH, Asfaw B, Manega P,
skelton of Homo habilis from Olduvai
Kaiser T, Bromage TG, Schrenk F (1995) Homin-
idal corridor research project update: new
Pliocene fossil localities at Lake Manyara
and putative oldest Early Stone Age occur-
rences at Laetoli (Upper Ndolanya Beds),
Keith A (1948) A new theory of human evolu-
tion. CA Watts & Company, London
Kimbel WH, Walter RC, Johanson DC, Red KE,
Aronson JL, Assefa Z, Marean CW, Eck GG,
Bobe R, Hovers E, Rak Y, Vondra C, Yemane
T, York D, Chen Y, Evensen NM, Smith PEB
(1996) Late Pliocene Homo and Oldowan
tools from the Hadar Formation (Kada
Hadar Member), Ethiopia. J Hum Evol 31:
549–562
King W (1864) The reputed fossil man of the
Neanderthal. Q J Sci 1: 88–97
Kullmer O, Sandrock O, Abel R, Schrenk F,
Bromage TG, Juwayyi Y (1999) The first
Paranthropus from the Malawi Rift. J Hum
Evol 37: 121–127
Leakey RE (1973a) Evidence for an advanced
plio-pleistocene hominid from East Rudolf,
Leakey RE (1973b) Further evidence of Lower
Pleistocene hominids from East Rudolf, north
Leakey LSB, Tobias PV, Napier JR (1964) A new
species of the genus Homo from Olduvai
Lewis OJ (1989) Functional morphology of the
evolving hand and foot. Oxford University
Press, Oxford
Linnaeus C (1758) Systema naturae. Laurentii
Salvii, Stockholm
Mayr E (1944) On the concepts and terminolo-
gy of vertical subspecies and species. Natl Res
Council Bull 262: 11–16
Prat S, Brugal J-P, Tiercelin J-J, Barrat J-A,
Boh M, Delagnes A, Harmand S, Kimeu K,
Kibunjia M, Texier P-J, Roche H (2005) First
occurrence of early Homo in the Nachukui
Formation (West Turkana, Kenya) at 2.3–2.4
(ed) Evolutionary history of the "Robust"
Australopithecines. Aldine de Gruyter, New
York, pp 383–403
Ramirez Rozzi F (1997) Les hominidés du Plio-
Pléistocène de la vallée de l’Omo. Cahiers de
paléoanthropologie CNRS éditions, Paris
Rayner RJ, Moon BP, Masters JC (1993) The
Makapansgat australopithecine environment.
J Hum Evol 24: 219–231
Roche H, et al. (2003) Les sites archéologi-
ques pliopléistocénes de la formation de
Nachukui, Ouest Turkana: bilan synthétique
Schoetensack O (1908) Der Unterkiefer des
Homo heidelbergensis aus den Sanden
von Mauer bei Heidelberg. W Engelmann,
Leipzig
Schrenk F, Bromage TG, Betzler C, Ring U, Juwayeyi Y (1993) Oldest Homo and Plio-
Stern JT, Susman RL (1983) Locomotor anato-
my of Australopithecus afarensis. Am J Phys
Anthropol 60: 279–317
Suwa G, White TD, Howell FC (1996) Mandib-
ular postcanine dentition from the Shungura
Formation, Ethiopia: crown morphology,
taxonomic allocations, and Plto-Pleistocene
hominid evolution. Am J Phys Anthropol
101: 247–282
new level of organization in cerebral evolu-
Tobias PV (1991) The skulls, endocasts and
teeth of Homo habilis. In: Olduvai Gorge,
vol 4, 2 vols. Cambridge University Press,
Cambridge
Tobias PV, von Koenigswald GHR (1964) A
comparison between the Olduvai hominines
and those of Java and some implications for
Toth N (1985) Archeological evidence for pref-
rential right-handedness in the lower and
middle Pleistocene, and its possible implica-
Vrba ES (1985) Ecological and adaptive changes
associated with early hominid evolution. In:
Alan R. Liss, New York, pp 63–71
Vrba ES (1988) Late Pliocene climatic events
and human evolution. In: Grine FE (ed)
Evolutionary history of the “Robust” Austra-
lopithecines. Aldine de Gruyter, New York,
pp 405–426
Vrba ES (1992) Mammals as a key to evolution-
ary theory. J Mamm 73: 1–28
Wood BA (1991) Koobi Fora research project,
vol 4: Hominid Cranial remains. Clarendon
Press, Oxford
Wood BA (1992) Origin and evolution of the
Wood BA, Collard M (2001) The meaning of
Homo. Ludus Vitalis vol. IX. 15: 63–74
10 Homo ergaster and Its Contemporaries

Ian Tattersall

Abstract

On the basis of their strong morphological differences from the Javan type materials, many authorities now consider the diverse East African fossils initially classified as “African Homo erectus” to be more properly allocable to the species H. ergaster. However, while this separation at the species level of the African and Indonesian hominids is certainly justified, the species H. ergaster as thus constituted still embraces a bewildering morphological variety. Indeed, although this grouping of African fossils seems to form a fairly coherent clade, it also appears quite diverse. The East Turkana type mandible of H. ergaster is matched by other specimens from Kenya and Tanzania, but not by the mandible of the iconic WT 15000 skeleton, and in its turn this specimen fails to match either in its cranial construction or its upper dentition most of the other comparable specimens usually referred to H. ergaster. Clearly there is a need for a systematic reappraisal of the entire “African Homo erectus” = Homo ergaster group; and equally clearly the hominid evolutionary story throughout the Old World in the Early–Middle Pleistocene was more complex than is implied by the extension of the species H. erectus to cover the entire miscellaneous assemblage of hominid fossils from this time period.

10.1 Introduction

There is probably no area of paleoanthropology in which disagreement is more profound than in the systematics and taxonomy of the genus Homo in the Early to Middle Pleistocene. This discord has a long and, dare one say it, illustrious pedigree, dating right back to the initial discovery and description of the Javan species Pithecanthropus (=Homo) erectus by Eugene Dubois in the early 1890s. At that time the only extinct hominid known was the European H. neanderthalensis, a form that, though peculiar in morphology, possessed a brain of modern human size. The new and more ancient hominid announced by
Dubois as an intermediate between modern humans and apes (a status reflected in his initial choice of name, which translates as “upright ape-man”) was thus the first known human fossil relative to display a brain cavity that was significantly larger than those of modern-day great apes while lying below the range of *H. sapiens*. Dubois’s discovery unleashed an immediate furor. The key to Dubois’s interpretation of this specimen as a human relative (though he stopped short of placing his find in the human family Hominidae) was the association of the type Trinil skullcap with a femur whose morphology was without doubt that of an upright biped in the modern fashion. This association was immediately attacked (and has continued to be periodically questioned), initially by those who preferred to see the cranium as that of a specialized ape, maybe related to the gibbons. At the same time, many of those who accepted the association between the cranium and femur wrote off the former as deriving from an aberrant modern human (see discussion in Tattersall 1995).

Still, some paleoanthropologists (Cunningham 1895) did seize immediately upon the Trinil specimen as an evolutionary intermediate between great apes and humans, and were willing to view the Javan hominid as an early member of a lineage that had given rise to *H. sapiens* via the Neanderthals. This interpretation rapidly gained ground (Theunissen 1988), and by early in the twentieth century, not least through the efforts of Schwalbe (1899)—and despite those of Boule (1911–1913)—the place of *H. erectus* as the “hominid in the middle” had effectively been secured. Given the tiny size of the hominid fossil record at that time, and that the apparent rudiments of a transformation series in brain size were present in what was known, this interpretation was hardly surprising: indeed, it was a good story that was hardly contradicted by the few facts then available. And, in the decades before the Second World War, two additional developments conspired to keep *H. erectus* at front and center in scenarios of human evolution.

The first of these was the discovery of the huge trove of *Sinanthropus pekinensis* fossils at Zhoukoudian near Beijing during the late 1920s and the 1930s, and of the similarly impressive series of *H. soloensis* crania at Ngandong in Java in 1931/1932. The hominids from both sites were reckoned to be very close to Javan *H. erectus*, if not exactly the same thing, and at a time when most hominid fossil sites produced a specimen here and there, both discoveries were overwhelming by virtue of the sheer volume of material produced. At the same time geneticists, systematists, and paleontologists in the USA and Europe were busily constructing the outlines of what came to be known as the Evolutionary Synthesis, which saw the gradual modification of continuous lineages as the central feature of the evolutionary process (see discussion in Tattersall 1995). And at mid-century, the ornithologist Ernst Mayr (1950), one of the principal
architects of the Synthesis, bluntly told the paleoanthropological profession that *S. pekinensis*, *H. soloensis*, and other Middle Pleistocene hominids all belonged to *H. erectus*, the species that occupied the middle part of a direct and gradually transforming lineage running from *H. transvaalensis* (the australopiths) at the beginning to *H. sapiens* (which embraced the Neanderthals) at the summit.

Mayr’s short article was perhaps the most influential contribution ever in paleoanthropology, and effectively set its agenda for the next half-century. The rapidly increasing size of the human fossil record eventually forced even Mayr to relent and to admit a little more complexity into the picture; but for decades paleoanthropologists labored steadfastly under the notion that the evolutionary history of our kind had largely involved the gradual modification through time of a central lineage that eventually culminated in *H. sapiens*. Of course, it was admitted that at any one point in time such a lineage, widely distributed across the Old World, would have harbored a variety of local variants; but throughout the second half of the twentieth century the emphasis was principally on within-species variation, rather than on the question of whether a signal of systematic (species) diversity might be detectable in the variety of morphologies that emerged as the hominid fossil record steadily enlarged. Against this background, the category *H. erectus* became a catchall for a huge and unwieldy assortment of fossils of substantially differing morphologies.

Such hominids came from widely scattered localities. First, the probably 1.0 to 0.7-million year (myr) old Trinil specimens from Java were joined by a steady stream of discoveries in the nearby Sangiran Dome, not far away, that probably date in the 1.5 to 1.0-Myr range, most of them closer to its younger end. Then the sample was augmented by the Chinese Peking Man fossils, now thought to be probably between about 500 and 300 thousand year (kyr) old, followed by the Ngandong specimens (which may be as young as 50 to 30 kyr old), and ultimately by other Javanese fossils from localities such as Sambung macan and Ngawi, both uncertainly dated but unlikely to be more than 200-kyr old and most probably younger. In China, later finds attributed to *H. erectus* came from sites including Lantian (Gongwangling and Chenjiawo, both perhaps around 1.0 Myr), Hexian (maybe 400 kyr), Nanjing (ca. 350 kyr), and even Longgupo, a site that may possibly be as much as 1.8-Myr old. Some European specimens in the 400 to 300 kyr range, such as those from Verteszöllös (Hungary), Arago (France), and Bilzingsleben (Germany), have been referred by some authors to *H. erectus*, as has the 900 to 800 kyr old calvaria from Ceprano in Italy. Further east, in the Caucasus, the 1.8-myr-old Georgian site of Dmanisi has yielded fossils that have been attributed to *H. erectus*. In Africa practically anything from the earlier Middle Pleistocene, and soon many older specimens as well, found themselves identified as *H. erectus*, so that the species came to
include such motley fossils as the 1.4 myr-old Olduvai Hominid (OH) 9 calvaria from Tanzania; the 700 kyr-old mandibles from Tighenif (Ternifine) in Algeria; the 400 kyr-old Salé partial braincase from Morocco; the 1.6 myr-old WT 15000 “Turkana Boy” skeleton from Nariokotome on the western side of Lake Turkana in Kenya, and several crania and mandibles in the 1.9–1.5 myr range from Koobi Fora and Ileret on the eastern side of the same lake; also from Kenya, the fragmentary Olorgesailie hominid at 1.0–0.9 myr; the 1 myr-old Daka calvaria from Ethiopia; a cranium of apparently similar age from Buia in Eritrea; and even, from South Africa, the perhaps 1.6 myr-old Swartkrans Member 1 SK 847 partial cranium.

Not only do these fossils cover an enormous span of time (ca. 1.8–0.03 myr) but they also embrace a huge range of morphologies, and taken together they hardly suggest a neat chronological series of the kind the Synthesis had predicted. Clearly, there is a systematic signal of some kind in the assemblage of hominid fossils that have at one time or another been allocated to this species: a signal of diversity at the species as well as the morphological level. But it was not until the beginning of the final quarter of the twentieth century that this possibility began to be seriously investigated.

10.2 Enter Homo ergaster

The existence of H. erectus as a convenient catchall for a remarkable variety of hominids certainly facilitated the telling of a relatively simple and straightforward human evolutionary story that could be told in terms of consistent long-term selection pressures for such things as more perfect thermoregulation, more efficient digestion, and above all greater intelligence. However, this story of the gradual honing over time of an ever more effective human machine was contradicted by the growing post-Synthesis realization that the evolutionary process consists of a great deal more than simple natural selection (Tattersall 1995, 1998). It also sat rather uneasily with the fact that the Pleistocene was increasingly being seen as a period of extraordinary short-term climatic oscillations as well as of longer-term fluctuations. And, most significantly of all, it was not the story that the growing assemblage of Homo fossils seemed to be telling.

The first shot across the bows of the all-encompassing notion of H. erectus came from work on fossils found at the classic African localities on the eastern shores of Lake Turkana. During the 1970s, several fossils, notably the partial crania KNM-ER 3733 and KNM-ER 3883, were discovered at Koobi Fora that their describers (Leakey and Walker 1976; Walker and Leakey 1978) ascribed to the species H. erectus. With cranial volumes of 848 and 804 ml, respectively (these and most of the other endocranial volumes cited here come from Holloway et al.
2004), these 1.8 and 1.6 myr-old individuals had possessed brains almost as large as that of Dubois’s much younger (1.0–0.7 myr) Trinil type specimen of *H. erectus*, at about 950 ml. Another significant specimen was a well-preserved mandible, KNM-ER 992 (Figure 10.1), about 1.5 myr old, which was recovered at Ileret and described simply as *Homo* of indeterminate species (Leakey 1972). Soon thereafter, Groves and Mazak (1975) jumped into the fray and made ER 992 the type specimen of a new species, *H. ergaster*. Although this innovation was disdainfully dismissed by the Koobi Fora team, and some other influential workers (Rightmire 1990) also continued to prefer the more comprehensive concept of *H. erectus*, this move finally opened the door to a reappraisal of the tradition of automatically assigning to *H. erectus* any and all African fossils with measured or assumed brain sizes in the general range of those noted above. In short, it became possible to entertain the notion that *H. erectus* is a terminal eastern Asian hominid species, and that hominid evolution throughout the Early and Middle Pleistocene continued in the pattern already established, with a vigorous exploration of the many different ways in which it was possible to be a hominid in the shifting and highly varied habitats of the Ice Age Old World. Section 10.3 will examine the morphological evidence for diversity within the immediate group to which both *H. ergaster* and *H. erectus* belong.
10.3 Cranial morphologies in *Homo* of the Early to Middle Pleistocene

### 10.3.1 General considerations

Nobody doubts that the *H. erectus/H. ergaster* group represents a relatively cohesive subset of the family Hominidae. The question, clearly, is whether in this group of fossils we are looking at a radiation of species, or at a single hugely variable species that may or may not have evolved directionally over the entire expanse of time (about 1.9–1.8 to 0.03 myr) and space (virtually the entire habitable Old World) it occupied. The distinction here is an important one, for species (even if not greatly differentiated morphologically from their closest relatives) are historically individuated entities which can compete with one another and become extinct without trace, whereas within species even demes that are significantly differentiated morphologically remain ephemeral entities that can disappear simply by absorption into ongoing conspecific populations. The twin processes of speciation and genetic/morphological differentiation are not linked, so that speciation may take place in the absence of significant morphological divergence, while the latter can occur without speciation intervening. This, of course, often makes unequivocal species recognition difficult in fossil assemblages (Tattersall 1986). However, it seems generally to be the case among living primates that, where substantial osteological differences are present among populations, those populations tend to act in nature as distinct species (i.e., as effectively independent reproductive entities). If we apply this criterion very conservatively to species recognition in the fossil record, demanding that the fossil species we recognize consistently bear distinctive osteological differences from related forms, we will probably underestimate the number of species in that record. However, we will not distort its overall pattern (Tattersall 1986, 1992). It is important to bear in mind that not all “morphs” (distinctive morphological entities) that we recognize in the fossil record will necessarily correspond to species in the reproductive sense; but it is equally evident that, given the nature of the fossil record, morphology must be the starting point in our analyses of it. After all, neither geological age nor geographical provenance, the other two attributes of any fossil, is *necessarily* linked to species identity, while its morphology is the only feature that makes a species recognizable at all. It is in this spirit that the remainder of this survey is offered, with the proviso that we will clearly not learn much that is useful about the pattern of events in hominid evolution during the Pleistocene if we do nothing more than replace the term “African *Homo erectus*” with the equally sweeping *H. ergaster*.
10.3.2 Eastern Asia

Given the established conventions of nomenclature and systematics, when we begin to consider the mass of material that has at one time or another been allocated to *H. erectus/H. ergaster*, we must, of course, begin with Dubois’s holotype material (Figure 10.2) from Trinil in Java (Schwartz and Tattersall 2003, 2005). It is the morphology of the Trinil skullcap that defines the species *H. erectus*, and the allocation of other fossils to this named entity must be done on the basis of their morphological similarities to it. The problem lies, of course, in deciding just how close those similarities should be, and it has to be admitted that there is no quantifiable answer to this question.

The Trinil 2 holotype is highly derived among hominids in a number of characteristics, especially of the brow region and the rear of the skull (Schwartz and Tattersall 2000, 2005). It is a smallish, long, thin-boned calotte with a narrow, shelf-like and laterally flaring postorbital region that flows onto the long, gently sloping, and flattish frontal plane with an almost imperceptible midline keel defined by shallow depressions bilaterally. The lateral walls of the low braincase are short and are markedly tilted inwardly above faint, low-set temporal lines; and the rather acute nuchal angle is distended posteriorly into a well-defined horizontal torus. Although the hominid sample from the adjacent Sangiran region is quite variable, especially in robusticity, the basic Trinil braincase morphology is repeated in most of the crania, the major exception being the quite complete if...
somewhat distorted cranium Sangiran 17 (Figure 10.3). The Sangiran sample of upper and lower dentitions is also heterogeneous, suggesting that a second hominid morph may be present in addition to the Trinil one.

The two (now three) crania known from the Sambungmacan region, to the northeast of Sangiran, are substantially younger than the Trinil/Sangiran assemblage and share a contrasting brow structure in which the quite horizontal supraorbital tori thicken laterally and appear to be continuous across glabella. The braincase itself has the appearance of being rather better inflated than typical of the Trinil form (the two published brain volumes are 1,035 and 917 ml), and the coronal profile is tent-shaped rather than having a squat and rounded outline. The nuchal plane undercuts the occiput to produce a horizontal torus that is well defined below, but has much poorer definition above. Together, these specimens, along with another calvaria from Ngawi (870 ml), produce a morph that generally resembles the Trinil type but is readily distinguishable from it. The Ngandong crania are similarly distinctive. Large, and more robust than the others, with endocranial volumes that range from 1,015 to 1,250 ml, they differ from the Trinil form in the ways in which the Sambungmacan ones do. However, in addition they present yet more capacious braincases that have more or less vertical side walls with quite aggressively raised temporal lines; and they show a greater rearward projection than the Sambungmacan forms do of the occipital torus, a feature with a well-defined superior border. Nonetheless, this entire group is united, in particular, by a set of derived supraorbital and nuchal morphologies, and it presents itself as a relatively cohesive whole. Clearly, these Javan forms are part of the same eastern Asian hominid clade and, if it is not divided up (basically, at this point, a matter of taste), it is this assemblage that must provide the core identity of the species *H. erectus*.

The hominid fossils from Locality 1 of Zhoukoudian, in China (Figure 10.3), have generally been considered classic exemplars of *H. erectus*. But it is still worthwhile noting that as a group they do differ fairly markedly from the Trinil type material, though mostly in ways that recall the Sambungmacan/Ngandong series. The crania from this site (with brain volumes ranging from 850 to 1,140 ml) are most distinctive in their supraorbital morphology, with a low-set glabella, supraorbital tori having a strong vertical component, and a continuous posttoral sulcus. In addition, there are marked dental differences between the Zhoukoudian and Sangiran samples (Schwartz and Tattersall 2005). Chinese specimens closely resembling the Peking Man materials include the Nanjing crania (Lü and the Tangshan Archaeological Team 1996). Other materials sometimes associated with them, such as those from Lantian, Hexian, Yunxian, and Longgupo, show a variety of differences both from Zhoukoudian and among themselves. These differences are discussed by Schwartz and Tattersall (2005).
Figure 10.3
Lateral views of crania from Africa and Eurasia in the “Homo erectus/Homo ergaster” group.
Left column, top to bottom: Sangiran S17; Koobi Fora KNM-ER 3733; Koobi Fora KNM-ER 3732; West Turkana KNM-WT 15000; Dmanisi D 2282. Right column, top to bottom: Zhoukoudian cranial reconstruction (Sawyer and Tattersall version); Koobi Fora KNM-ER 3883; Olduvai OH 9; Koobi Fora KNM-ER 1813 (scale = 1 cm). From Schwartz and Tattersall (2005). Photographs by and © Jeffrey H. Schwartz; Turkana fossil images courtesy of National Museums of Kenya.
10.3.3 Africa

As already noted, at one time or another many African fossils in the 1- to 2-myr time range have been referred to the species *H. erectus*. Among them, the classic exemplars are fossils from the Turkana Basin of northern Kenya, notable among these being the cranium KNM-ER 3733 and the calvaria ER 3883 from sediments at Koobi Fora to the east of Lake Turkana, and the fairly complete skeleton KNM-WT 15000 from deposits to its west at Nariokotome. The mandible KNM-ER 992, initially allocated simply to *Homo* sp., comes from Ileret, to the north of Koobi Fora. All of the specimens concerned date within the 1.9- to 1.5-myr range. The widely used term “African *Homo erectus*” was a convenient designation for these fossils and a host of others, but it disguises the fact that a substantial variety of morphologies is involved.

This reality was first acknowledged in 1975 by Groves and Mazak, who as noted made the ER 992 mandible (Figures 10.1 and 10.4), the holotype of the new species *H. ergaster* (“work man”). Subsequently, many authors have begun to use the new name in place of “African *Homo erectus*” to the extent that it is now *H. ergaster* that is the standard-issue *Homo* of the 2 to 1 myr period. Still, all this move has achieved is to remove the assortment of Asian morphologies from the African equation, and it does nothing to address the morphological variety found within the continent in this period. In coming to grips with this, the best place to start is with the iconic example, the KNM-ER 3733 cranium (Figure 10.3), which has an endocranial capacity of 848 ml. In this individual, the supraorbitals arc separately over each orbit and project forward as well as upward. There is thus a distinct posttoral sulcus in front of the quite steep frontal rise, which rapidly peaks before the profile descends more gradually rearward. Seen from behind, the braincase is rather tall compared to its breadth, and its side walls are curving. The raised temporal lines start quite far medially. It is unsurprising that these characteristics distinguish this specimen sharply from any Asian *Homo*; more remarkable are its differences from the ER 3883 cranium (Figure 10.3), which has an endocranial volume of 804 ml. This individual has very thickened supraorbital margins that protrude outward but slightly down, overhanging nasion and the (mostly missing) face beneath. Further, in this specimen the frontal slopes strongly up and back, reaching its maximum height rather far back. Unlike in ER 3733, the mastoid is large and protruding, and in what is preserved of the face the zygoma flares outward from top to bottom. That this morphology is no freak is shown by its close repetition in preserved features of the ER 3732 partial cranium (Figure 10.3).

Interestingly, all of the specimens just mentioned are distinctly different from the skull of the KNM-WT 15000 skeleton (Figure 10.3), exhaustively
described in the monograph edited by Alan Walker and Richard Leakey (1993). This 1.6 myr-old skeleton is remarkable both for its degree of completeness and for being the earliest good evidence we have in the human fossil record of the arrival of essentially modern stature and postcranial proportions. Frustratingly,

Figure 10-4

Homo ergaster and its contemporaries
we cannot yet be certain that this was the case for the possessors of the ER 3733 and ER 3883 crania. The adolescent WT 15000 individual died at about 8 years of age, but was at a stage of development approximating that of a modern 12-year-old. This presumed male stood about 160 cm tall, but had he survived to adulthood it is estimated that he would have topped 180 cm, and he was long-limbed and slender, with efficient heat-shedding proportions that would have served him well in the heat of the open tropical savanna. In contrast to the relatively long crania just described, the braincase of WT 15000, with a capacity of about 900 ml, was quite short and had a well-rounded profile. To the extent that it is possible to judge, the badly damaged supraorbital surfaces amounted to little more than substantial thickenings of the superior orbital margins, lacking either the aggressive projection seen in ER 3733 or the vertical thickening noted in ER 3883. The structure of the face contrasts with that seen in both ER 3733 and ER 3883; it is longer and narrower, with much more alveolar prognathism, a higher and narrower nasal aperture, and preserved portions of the nasal bones that suggest a flatter profile of the upper face. It is often claimed that these differences from the east Turkana specimens are due to the subadult status of WT 15000, but this appears rather dubious since differences of this kind would, if anything, probably have become more marked with age.

Cranial differences are backed up by dental comparisons to the extent that these are possible. ER 3883 has no associated teeth, and ER 3733 has only one, an upper M2. But this tooth, though unfortunately quite heavily worn, is nonetheless distinctly different from its counterpart in WT 15000. The upper M2 of ER 3733 has smooth enamel and well-defined trigon cusps, with the paracone much larger than the metacone. The hypocone is distally placed, the cristae are sharp, and both basins are well excavated. In WT 15000, in contrast, the upper M2s are high-crowned with fairly flat but wrinkled occlusal surfaces. The cusps of the trigon are subequal in size, and the basins are quite shallow. In both ER 3733 and WT 15000, the upper M2s contrast with their homologues in the best-preserved upper dentition from Java, the Sangiran 4 palate, in which there are low cusps, a massive hypocone, and a very large postprotocrista that is not a feature of either African specimen.

The ER 992 type specimen of *H. ergaster* is a lower jaw, so comparisons to ER 3733 and ER 3883 are not possible. However, uniquely, the WT 15000 skull has a definitively associated lower jaw (Figure 10.4), allowing direct comparison to ER 992. When this comparison is made, clear differences become apparent. In ER 992 (Figure 10.4) the lower canines are quite high and are compressed buccolingually. In the anterior lower premolar there are distinct anterior and posterior foveae, and the protoconid is the clearly dominant cusp. In the
posterior premolar the protoconid and metaconid are subequal, and the basins are shallow. The elongated lower molars bear rounded and protruding hypoco- nulids, their basins are shallow, and their enamel is wrinkled. In contrast, the lower canines of WT 15000 are short-crowned, with distinct mesial and distal foveae that bound a strong lingual pillar that swells out the tooth at its base. Both premolars have deep mesial and distal basins. The first molar is distended mesially, and both erupted molars have large and lingually placed hypoconulids and narrow but deep talonid basins that are surrounded by well defined but rather bulbous cusps. Here, again, we have two distinctly different lower dental morphologies, both of which also differ from their homologues known from Sangiran. Morphologically, at least, the lower dentitions of ER 992 and WT 15000 are not “the same thing,” and both are at variance with Javan H. erectus (Schwartz and Tattersall 2000, 2005).

Interestingly, the 1.5-Myr-old ER 992 from Ileret makes a fairly good match for the mandible OH 22 (Figure 10.4) from Olduvai Gorge, in Tanzania, as well as for its rather older (1.9 Myr) neighbor ER 3734 (Figure 10.4) from Koobi Fora. As for WT 15000, its lower teeth compare quite closely with those of OH 13, one of the paratypes of H. habilis from Bed II of Olduvai Gorge and of similar age. Both show mesially tapering premolars, with small metaconids lying opposite the protoconids, and smaller foveas in front of these cusps than behind. The degree of wear on the molars is very different, but both show an oblique groove that runs between the hypoconulid and hypoconid to the base of the metaconid; and in both the hypoconid lies buccally and the M1s taper slightly distally while the M2s are more broadly rounded at the rear. In the cranium and upper dentition, WT 15000 shows substantial similarities with the east Turkana cranium ER 1813 (Figure 10.3). Although the latter boasts a substantially smaller intracranial volume of 510 ml, both specimens share a short, high cranial vault (rather like that of OH 13) with rounded brows that arc over each orbit and a frontal that rises behind a very short posttoral plane. In both, the nasal apertures are tall, relatively narrow and taper strongly upward, and the nasoalveolar clivuses are long and slope forward.

Even more telling are upper dental similarities among WT 15000, ER 1813, and OH 13. All have a high-crowned but flat-surfaced M1, with the mesiodistally long hypocone as high as the protocone and separated from it by a lingual notch. In all, the M′s are basically similar to the M1s, but show reduction of the metacone and a smaller notch between the hypocone and protocone. In all, M1 and M2 both have thick postcingula. In WT 15000 and ER 1813, the P2′s are very similar in having a bulbous and centrally placed paracone and a continuous crista running mesially from the paracone and swinging right around the side of the
tooth. Further, although it has a very worn dentition the recently described palate OH 62 seems to present an upper dental morphology similar to those of the three specimens just described. In sum, the evidence seems to be quite compelling for the existence of an upper dental morph, most spectacularly represented by the fairly complete individual WT 15000, that is found almost a thousand kilometres away at Olduvai as well as around Lake Turkana at Koobi Fora and Nariokotome. The available name for this morph, should anyone wish to designate it a species, is *H. microcranous* (Ferguson 1995).

Other “African Homo erectus” specimens include the 1.4 myr-old OH 9 calvaria (Figure 10.3: 967 ml) from Olduvai Gorge, but it does not compare any better to material from eastern Asia than it does to the Turkana fossils, and though it has been compared to the purportedly *H. erectus* Daka cranium from Ethiopia (~1.0 Myr, 995 ml; Asfaw et al. 2002), resemblances between the two other than in endocranial volume and its correlates are not particularly striking (Schwartz and Tattersall 2005). In sum, there is considerable morphological diversity among African *Homo* of the 2.0 to 1.0 myr period; and this diversity does seem to be organized into a number of distinctive morphs. Some of these are represented by individual fossils such as OH 9 that are clear outliers in terms of other known material; others seem to be represented by multiple individuals and even at multiple sites. Definitive systematic organization of this variety will clearly have to await a more comprehensive fossil record, but it is already evident that we are not looking here at a chronological transformation series, even one that is represented by high diversity at all time points. Somewhere in all of this there is a systematic signal, and it is evident that the blanket appellation *H. ergaster*, while a useful device for distinguishing the African hominid radiation of this period from the Asian one, is an inadequate expedient for describing diversity in the African record.

### 10.3.4 Europe

It is now fairly widely accepted that most of the western European forms that have at one time or another been described as *H. erectus* (see above) are better allocated to *H. heidelbergensis*, a move that takes these fossils out of the scope of the current discussion. And while some still view the Italian Ceprano calvaria (ca. 800–900 kyr; 1,165 ml), recently designated the holotype of *H. cepranensis* by Mallegni et al. (2003), as a representative of *H. erectus*, nobody has claimed that it represents *H. ergaster*. Indeed, its closest comparison (Schwartz and Tattersall 2005) seems to be with the Ethiopian Daka specimen with which, as discussed above, there would be little sense in trying to associate the Ileret holotype.
To the east, however, in the Caucasus at the Georgian site of Dmanisi (1.8 myr), there exists a very important and early hominid fossil assemblage with claimed African affinities (Figures 10.3–10.7). The first find, the mandible D 211 (Figure 10.4) discovered in 1991, was assigned by its describers to “archaic African Homo erectus” (Gabunia and Vekua 1995). It was more generally attributed to H. erectus by Henke (1995), and also by Bräuer and Schultz (1996)—although these latter authors remarked that, oddly, this early mandible showed “progressive” features seen in geologically younger H. erectus. These early differences in interpretation foreshadowed a fairly wild taxonomic ride. Gabunia et al. (2000) reported the discovery of two crania (D 2280 and 2282; Figure 10.3) close to the original site; they considered these comparable in size and morphology to Koobi Fora H. ergaster, although cranial volumes were somewhat smaller: 780 and 650 ml, respectively. With the discovery in September 2000 of a very large and long mandible with highly worn teeth (D 2600; Figure 10.5), the

Figure 10.5
The D 2600 mandible from Dmanisi. (a) Front view; (b) left lateral; (c) occlusal. Courtesy of David Lordkipanidze
picture changed again. This mandible presented a marked contrast to D 211, but the Dmanisi team nonetheless concluded that all of the specimens belonged to a single highly sexually dimorphic species which they named *H. georgicus*, with D 2600 as its holotype (Gabounia et al. 2002). The gracile specimens D 211, 2280, and 2282 were considered to be female; and the robust D 2600 lower jaw, plus material yet to be described, was viewed as a male representative of this species. The group concluded that *H. georgicus* “preserves several affinities with *Homo habilis* and *Homo rudolfensis* . . . foretelling the emergence of *Homo ergaster*” (Gabounia et al. 2002 p 245). But things did not stop there. Almost simultaneously, the team announced the discovery of an associated cranium and mandible, D 2700/D 2735 (Vekua et al. 2002; Figure 10.6). Strikingly different from the crania discovered earlier, though also notably small-brained (600 ml), this excellently preserved specimen was said by its describers to bear resemblances to penecontemporaneous East African fossils. Abandoning the species *H. georgicus*, as well as the notion of *H. ergaster* as a separate entity, the Dmanisi team allocated D 2700/D 2735, and the rest of the hominid assemblage along with it, to *H. erectus*, while noting that they “are among the most primitive individuals so far attributed” to that expanded species (Vekua et al. 2002 p 88).

During the 2002/2004 field seasons yet another associated cranium and mandible (D 3444/D 3900) were recovered at Dmanisi. The most remarkable aspect of this aged presumed male (Figure 10.7) is that he had possessed just a single tooth at death and had evidently been largely edentulous for many years (Lordkipanidze et al. 2005). Although at least one recent chimpanzee is known to have survived a long time in an edentulous state, the Dmanisi team surmised that the individual must have “survived for a lengthy period without consuming foods that required heavy chewing . . . and/or by virtue of help from other individuals” (Lordkipanidze et al. 2005 p 718), and suggested that this had significant implications for early hominid social structure. They also noted that the cranium had been found in close proximity to Mode 1 stone artifacts and to cut-marked animal bones. The authors refrained from commenting on the systematic implications of the new find; but in a review published soon afterward, Rightmire et al. (2006) reaffirmed their belief that the Dmanisi assemblage as a whole was a single “paleodeme” best placed within *H. erectus* (which to them subsumed *H. ergaster*), despite resemblances to *H. habilis* in brain volume and in some aspects of craniofacial morphology. This notion may require reevaluation following the reported discovery of a cranium, yet to be described, that matches the long and robust lower jaw D 2600; but there is already a chink in the armor. For while Rightmire et al. (2006 p 140) noted that if D 2600 could be accommodated within the rest of the Dmanisi hominid population, then “the appropriate nomen is *H. erectus georgicus*,” they also pointed out that should the separate species status
advocated by Gabounia et al. (2002) for the large jaw be “verified by new discoveries, then a subspecies other than *H. erectus georgicus* will have to be selected [for the remainder of the sample].”

The taxonomic journey of the Dmanisi hominids reflects the unusual morphologies that make it hard to fit them into established categories. And this
Figure 10.7
journey is certainly not yet at an end. Schwartz and Tattersall (2005) have already noted that the morphological heterogeneity in the Dmanisi assemblage makes it difficult to recognize a single consistent morph at the site, irrespective of what this might imply about species status. Additionally, in their view none of this material appears to bear very close, i.e., systematically suggestive, resemblances either to any Asian fossils that have been described as *H. erectus*, or to any African specimens that have been allocated to *H. erectus* or to *H. ergaster*. Exactly how much systematic variety there is in this assemblage awaits more study; but although the Dmanisi hominids most plausibly represent one or possibly more early departures from Africa, hard on the heels of the origin of *Homo* as (probably) best defined by essentially modern postcranial form (though this remains to be tested by analysis of as-yet unpublished postcranial remains from Dmanisi), it is hard at present to point at craniodental morphologies that specifically unite them with any latest Pliocene or earliest Pleistocene African hominids yet known. And their allocation to *H. erectus* is, of course, only possible in the context of an inclusive view that subsumes all the very many diverse morphologies among the hominid fossils discussed in this chapter within a single ecumenical but morphologically meaningless species.

10.4 Conclusions

*H. ergaster* is the designation of choice for the growing number of paleoanthropologists who believe that the fossils previously allocated to “African *Homo erectus*” are sufficiently different from the Asian type material of *H. erectus* to warrant assignment to a distinct species. Adoption of this nomenclature is a considerable improvement on our older understanding, certainly to the extent that it emphasizes that the Trinil fossil and others like it are quite highly autapomorphic, and that Javan or at least eastern Asian *H. erectus* is thus most appropriately viewed as an indigenous and terminal regional species (or maybe even clade), rather than as an Old World–wide stage or grade in hominin evolution. Nonetheless, it remains true that to call the entire African rump of this Old World “*Homo erectus* grade” *H. ergaster*, is to brush a huge diversity of morphologies under the rug of one single species. Close examination of the morphologies displayed by the diversity of fossils that have at one time or another been referred to as “African *Homo erectus*” makes it evident that, while it is likely that all may be legitimately regarded as members of a single hominin clade, this clade appears to be fairly diverse.

The KNM-ER 992 holotype mandible of *H. ergaster* appears to be matched morphologically by the OH 22 mandible from Olduvai as well as by another
mandible (ER 3734) from Koobi Fora. But there seems to be no compelling reason to match these lower jaws with any of the cranial materials available, and certainly not to associate them with the iconic KNM-WT 15000 skeleton, the lower dentition of which is distinctly different in a whole host of characteristics. Clearly, there is a need for a detailed systematic reappraisal of the entire “African Homo erectus” = Homo ergaster group. Meanwhile, the recognition of a distinct H. ergaster clade at least serves to highlight the fact that the complexity of the hominin evolutionary story throughout the Old World in the Early-Middle Pleistocene was far greater than is implied by the inclusion of the entire group of fossils involved within the single hugely variable species H. erectus.

Acknowledgments

My appreciation goes to Jeff Schwartz, in collaboration with whom many of the notions advanced here were developed, and who kindly provided Figures 10.2–10.4. Ken Mowbray furnished the scan reproduced in Figure 10.1. David Lordkipanize kindly gave permission for the images reproduced in Figures 10.5–10.7, and Thorolf Hardt prepared them for publication here. Thank you all.

References

Cunningham D (1895) The place of “Pithecanthropus” on the genealogical tree. Nature 53: 269
the 10th international symposium on dental morphology. Christine und Michael Brünae GbA, Berlin, pp 459–464


Lù Z and the Tangshan Archaeological Team (1996) Locality of the Nanjing man fossils. Cultural Relics Publishing House, Beijing


Schwalbe G (1899) *Studien über Pithecanthropus erectus* Dubois. Morphol Anthropol 1: 16–228


11 Defining *Homo erectus*: Size Considered

**Susan C. Antón · Fred Spoor · Connie D. Fellmann · Carl C. Swisher III**

**Abstract**

*Homo erectus*: for some a single, widely dispersed, polytypic species ultimately ancestral to all later *Homo*, for others a regional, Asian isolate, a sidebranch of later hominin evolution. In some views, the definition of *H. erectus* expands to include all Early and Middle Pleistocene fossils from Africa, Europe, and Asia, whereas other views exclude the European fossils and still others include only portions of the Asian fossil record. Temporally, *H. erectus* may thus span from 1.8 to perhaps 0.025 Ma. In this chapter, we discuss the importance of body and brain size in the characterization of *H. erectus*. We also provide a historical review of the recovery of the fossil record of *H. erectus* because the understanding of its body size, the relationship between body and brain size, and the importance of the influence of scale on the cranial characters used in defining the taxon are all issues whose interpretation was in large part determined by the order and context in which fossil remains were recovered. We find evidence of clinal variation in stature in *H. erectus*, and across the taxon body size (stature) does not increase with time, whereas brain size gradually increases at a rate of about 160 cm$^3$/Myr. Cranial characters, particularly those related to vault thickness and development of the supraorbital torus and many of those related to differentiating African from Asian *H. erectus*, scale with brain size in *H. erectus* yielding little support for a differentiation between *H. erectus* and *H. ergaster*. In contrast, *H. erectus* is clearly differentiated from *H. sapiens* in terms of the scaling relationship between brain and body size and between brain size and several cranial measurements, including foramen magnum length. On these same relationships, it is difficult to differentiate the smallest *H. erectus* from the largest *H. habilis*.

### 11.1 Introduction

The definition of *Homo erectus* includes, in some views, all Early and Middle Pleistocene fossils from Africa, Europe, and Asia, whereas other views exclude the European fossils and still others include only the Asian or portions of the Asian
fossil record (Rightmire 1993; Wood 1994; Schwartz and Tattersall 2000; Asfaw et al. 2002; Antón 2003). There is also a contingent of workers who argue that *H. erectus* forms part of a single evolving lineage ending in *H. sapiens* and thus recognize only one species of hominin after about 1.8 Ma (Jelinek 1981; Aguirre 1994; Wolpoff 1999; Curnoe and Thorne 2003). For most who recognize the species, *H. erectus* is either a single, widely dispersed, polytypic species ultimately ancestral to all later *Homo* or a regional isolate restricted to Asia, a sidebranch of later hominin evolution. In all but the most speciose cases, *H. erectus* is likely to span from 1.8 to perhaps 0.025 Ma.

Here we discuss the importance of body and brain size in the characterization of this species and provide a historical review of the recovery of the fossil record of *H. erectus* in order to consider how this history shaped the interpretation of the taxon. Our understanding of body size in *H. erectus*, the relationship between body and brain size, and the importance of the influence of scale on the cranial characters used in defining the taxon are all issues whose interpretation was in large part determined by the historical order and context in which fossil remains were recovered.

## 11.2 *H. erectus* in historical context

Defining the taxon *H. erectus* has been an accretionary process over a century long. The view of the taxon has vacillated between the vision of *H. erectus* as more apelike or as more humanlike. The order of discovery of particular fossils has been, in part, random and their interpretation colored by other, previously known finds. Thus, it is useful to consider the order of the finds and the context in which their morphology was reviewed and the species definitions came to take shape. We divide this history into three periods, although they could be easily subdivided into several more. These are the pre-1950 period, 1950 to the mid-1980s, and 1990s to present (Figure 11.1).

### 11.2.1 The early discoveries

The earliest period of discoveries focused entirely on Asia. The species was, of course, originally dubbed *Pithecanthropus erectus*, by Dubois (1894) with the Trinil 2 calotte from Java serving as the type specimen. Aside from *P. erectus*, Neandertals were the only other fossil humans known and the initial challenge lay in demonstrating that the Indonesian fossils were, in fact, hominins (Dubois 1894, 1924, 1926a, b, 1932; Schwalbe 1899). The “discovery” of *Eoanthropus*
dawsoni (aka Piltdown Man) in England in the early 1900s did not help this case (Dawson and Smith Woodward 1913, 1914). Piltdown had a large brain and ape mandible, fitting the conventional wisdom of the time for the order of appearance of evolutionary adaptations in the hominin lineage. \( P. \text{erectus} \) did not. Although Dubois considered \( P. \text{erectus} \) directly ancestral to modern humans, other naturalists remained unconvinced (Figure 11.2).

Figure 11.1
History of major cranial discoveries of \( H. \text{erectus} \) by year of discovery and geological age. Note that the earliest discoveries incorporated the entire known time range of \( H. \text{erectus sensu lato} \). Diamonds indicate \( H. \text{erectus} \) from the following regions: closed diamonds, Indonesia; open diamonds, China; open circles, Africa; closed circles, Dmanisi, Georgia.

Figure 11.2
The position of \( P. \text{erectus} \) when first proposed. Either viewed as a great ape (left) or as a side branch of human evolution (right). Dubois, however, considered \( P. \text{erectus} \) directly ancestral to genus \( Homo \).
The late 1920s and 1930s yielded a wealth of Asian hominins that were initially assigned to several genera and species (Table 11.1). Excavations outside Beijing at Zhoukoudian, China between 1928 and 1936 yielded relatively complete calvarial remains that were assigned to the new genus and species, *Sinanthropus pekinensis* (Black 1931; Weidenreich 1936, 1937, 1941, 1943). Along the Solo River in Java, calvarial remains of 12 individuals were recovered between 1931 and 1933 in excavations at Ngandong and assigned to a new species *Homo*

Table 11.1  
Major *H. erectus* find sites and the original taxonomic attribution

<table>
<thead>
<tr>
<th>Site</th>
<th>Year of hominins</th>
<th>Taxonomic name</th>
<th>Geological age</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trinil</td>
<td>1890s</td>
<td><em>Anthropopithecus erectus</em></td>
<td>0.9 Ma</td>
</tr>
<tr>
<td>Kedungbrubus</td>
<td>1890s</td>
<td><em>Pithecanthropus erectus</em></td>
<td>1.5 Ma</td>
</tr>
<tr>
<td>Zhoukoudian</td>
<td>1928–1936, present</td>
<td><em>Sinanthropus pekinensis</em></td>
<td>400–600 ka</td>
</tr>
<tr>
<td>Ngandong</td>
<td>1931–1933, 1970s, present</td>
<td><em>H. (javanthropus) soloensis</em></td>
<td>50–27 ka</td>
</tr>
<tr>
<td>Perning</td>
<td>1936</td>
<td><em>H. (pithecanthropus) modjokertensis</em></td>
<td>1.8 Ma</td>
</tr>
<tr>
<td>Sangiran</td>
<td>1937 to present</td>
<td><em>P. erectus, P. dubius, P. robustus, P. modjokertensis Meganthropus paleojavanicus</em></td>
<td>1.66–1.0 Ma</td>
</tr>
<tr>
<td>Gongwangling</td>
<td>1960s</td>
<td><em>S. lantianensis cf. P. robustus</em></td>
<td>1.2 Ma</td>
</tr>
<tr>
<td>Sambungmacan</td>
<td>1973–present</td>
<td><em>H. erectus</em></td>
<td>Similar to Ngandong</td>
</tr>
<tr>
<td>Hexian</td>
<td>1980</td>
<td><em>H. erectus pekinensis</em></td>
<td>0.23 Ma</td>
</tr>
<tr>
<td>Africa</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ternifine</td>
<td>1950s</td>
<td><em>Atlanthropus mauritanicus</em></td>
<td>400 ka</td>
</tr>
<tr>
<td>Olduvai</td>
<td>1960s</td>
<td><em>H. leakeyi (OH 9) H. erectus</em></td>
<td>1.47–~1.0 Ma</td>
</tr>
<tr>
<td>Koobi Fora</td>
<td>1978–present</td>
<td><em>H. erectus/ergaster</em></td>
<td>1.9–1.5 Ma</td>
</tr>
<tr>
<td>Nariokotome</td>
<td>1984</td>
<td><em>H. erectus/ergaster</em></td>
<td>1.5 Ma</td>
</tr>
<tr>
<td>Danakil (Buia)</td>
<td>1995–97</td>
<td><em>H. sp.</em></td>
<td>1 Ma</td>
</tr>
<tr>
<td>Daka (Bouri)</td>
<td>1997</td>
<td><em>H. erectus</em></td>
<td>1 Ma</td>
</tr>
<tr>
<td>Olorgesailie</td>
<td>2004</td>
<td><em>H. erectus</em></td>
<td>0.9 Ma</td>
</tr>
<tr>
<td>Eurasia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ubeidiya</td>
<td>1959</td>
<td><em>H. sp.</em></td>
<td>1.4 Ma</td>
</tr>
<tr>
<td>Dmanisi</td>
<td>1990s–present</td>
<td><em>H. erectus/ergaster</em></td>
<td>1.7 Ma</td>
</tr>
</tbody>
</table>
(Javanthropus) soloensis (Openoorth 1932, 1937). In 1936, an immature fossil calvaria from Mojokerto was assigned to *Pithecanthropus modjokertensis* (von Koenigswald 1936). In 1937 and 1938, several calvaria and mandibular remains were recovered from the Sangiran Dome region of Java and assigned variously to *Pithecanthropus erectus, P. dubius, P. robustus, P. modjokertensis*, and *Meganthropus paleojavanicus* (von Koenigswald and Weidenreich 1938; von Koenigswald 1940).

Not surprisingly, it is impossible to identify a single, “accepted” view of the phylogenetic position of these fossils during this pre-1950s period. However, it should be recognized that the many generic and specific names assigned to the remains did not necessarily indicate that scientists of the time considered these fossils taxonomically distinct. Indeed, some of those most familiar with the anatomy of the fossils, especially Weidenreich, were clear that they considered the Chinese and Javanese remains to be geographic variants of the same taxon despite continuing the use of different specific and generic names (Dubois 1936; von Koenigswald and Weidenreich 1939; Weidenreich 1940, 1943, 1951). Thus, although Mayr (1944, 1950) is credited with synonymizing the multiple Asian taxa (and many African taxa) into *Homo*, and ultimately *H. erectus*, the scientific basis for this was explicitly laid out by the paleoanthropologists of the day.

There are two important attributes of this first group of Asian fossils. First, based on geological evidence, the fossils span from the earliest to the latest Pleistocene (von Koenigswald 1936, 1962; de Terra 1943). That is, the first period of discovery fortuitously sampled the entire known time range of *H. erectus sensu lato*. Although the extent of this Asian time range would be contracted by scientific thinking during the next period of discovery (Pope 1983), the early stratigraphic estimates of fossil age have been more recently confirmed by radiometric age estimates to encompass a range from about 1.8 Ma to perhaps 25 ka (Swisher et al. 1994, 1996; Larick et al. 2001). Second, despite this large time span, the cranial remains from Asia share a common morphological “bauplan.” The similarities among the Asian fossils are easily recognized as phylogenetic links through time, although geographic variation within the region is also well established in the sample (Weidenreich 1943; Antón 2002a; Kidder and Durband 2004). However, as might be predicted, the variation does not encompass all the variation of the fossil sample as known today.

A definition based on just these pre-1950s remains would recognize a moderately large brained (*x* = 1,000 cm³), thickly walled, angularly vaulted hominin with a series of cranial superstructures, such as sagittal keels, angular and occipital tori, and some cranial base differences from modern humans, including angulation of the petrous portion of the temporal. At that time brain size, but not body
size, was a point of interest probably because postcranial remains were scarce and because issues of the evolution of body size in the human lineage as a whole had yet to be addressed.

### 11.2.2 The 1950s through the 1980s

In the second period of discovery (1950s–1980s), African finds dominate the *H. erectus* discussion, and early in this period new finds continue to be assigned to new taxa. The Ternifine remains from North Africa are assigned to *Atlanthropus mauritanicus* (Arambourg 1954), and Olduvai Hominid 9 is initially assigned to *H. leakeyi* (Heberer 1963). Both, however, are quickly and easily subsumed into *H. erectus* (LeGros Clark 1964), as are postcranial remains from Olduvai Gorge (Day 1971), although whether this is the appropriate place for the North African fossils remains a question (Antón 2003). This tendency to lump all finds into *H. erectus* continues with the discovery in the late 1970s of more complete crania and cranial fragments from the Koobi Fora region of Kenya. In particular, KNM-ER 3733 and 3883, both of which retain substantial portions of the facial skeleton (Leakey and Walker 1985), are included directly into *H. erectus* (Walker 1981) and, more importantly, become iconic of the species. The fragmentary OH 12 is also recognized as *H. erectus* (Rightmire 1979) as are many fragmentary remains from Koobi Fora (Walker 1981; Walker et al. 1982). And the subadult skeleton KNM-WT 15000 is included in *H. erectus* (Brown et al. 1985) and becomes iconic for body size in *H. erectus*. There are also numerous Asian discoveries during this period, but they are not as systematically published as the African finds and do not depart markedly from the earlier Asian sample (Sartono 1964, 1981; Jacob 1966, 1984; Wu et al. 1966; Chiu et al. 1973; Wu 1983).

As a group, the African fossils have some characteristics that differ in nature from the earlier Asian discoveries. Importantly, some of the African remains retained relatively intact facial skeletons and thus became “the face” of *H. erectus*. The average cranial capacity was smaller than that of the Asian sample ($x = 870$ versus $1,025$; Figure 11.3), although within the Asian range. And the African fossils encompassed only the geologically older half of the *H. erectus* time range; most are 1.5 Ma and older, with none of the principal remains being much younger than 1.0 Ma (Figure 11.1). Despite these differences and the fact that Asian *H. erectus* fossils are more numerous, the African fossils came to dominate the discussion of *H. erectus*. This may in part be because individual African fossils are more complete and, as a group, more accessible. Indeed, fragmentary African remains, such as OH 12, were also largely ignored in discussions about the attributes of the species (Antón 2004).
The tendency to lump the African fossils into *H. erectus* suggests that through the early to mid-1980s *H. erectus* was often viewed as a single, widely dispersed, geologically long-lived, polytypic species, including both African and Asian fossils (Howell 1978; Howells 1980). However, the growing interest in African *H. erectus* also led to comparisons with “Asian *H. erectus (H. erectus sensu stricto)*” and often with the most easily accessible subset of these specimens, those from Zhoukoudian (Antón 2002a). Such work compared groups across not only geographic space (Asia versus Africa) but also across about 1 Myr of time (1.5 Ma and earlier for Africa and 0.5 Ma and later for China). These comparisons also tended to artificially reduce the amount of variation in the Asian fossil record. The morphological differences between the groups were interpreted in some cases as evidence that more than one species was included in *H. erectus sensu lato* (Andrews 1984; Stringer 1984; Wood 1984, 1991; Tattersall 1986). This trend of splitting African and Asian *H. erectus* corresponds historically with the use of cladistics becoming mainstream in paleoanthropology (Delson et al. 1977). In this context, it was perceived that only the Asian subsample had autapomorphic features. As we have seen in previous chapters of this volume, early African *H. erectus* would be called *H. ergaster* based on the name given to the type mandible KNM-ER 992 (Groves and Mazek 1975). Others interpreted differences between African and Asian samples as the result of geographic variation in a polytypic species (Rightmire 1990).

The crux of the argument over whether Asian and African *H. erectus* represent one or more species would rest on discrete characters said to differ between

---

**Figure 11.3**

Cranial capacity of *H. erectus* by year of discovery. Note that the type specimen discovered in the 1890s has about the mean cranial capacity for the species and that more recent finds have extended the lower limit of cranial capacity in the taxon.
the two regions. The argument entailed both the distribution of these characters and how one feels about geographic variation, with temporal and idiosyncratic variation playing little role in the debate. The features in question rest heavily on the greater expression in Asia than in Africa of cranial superstructures (e.g., bregmatic and metopic eminences, sagittal keel, angular and occipital tori), and generally thick cranial walls, large, continuous barlike supraorbital tori, lower and more angular cranial vaults. Some of these are, however, present in some of the African fossils (Bräuer and Mbua 1992). A bregmatic eminence appears in OH 12 and KNM-ER 3733; an angular torus appears in OH 9; an occipital torus restricted to the middle third of the occipital appears in most African H. erectus. Some characters, such as vault angulation, have been shown to vary similarly in both samples (Antón 2002a). Some differences may be due to the retention of primitive characters in the early African sample (e.g., anterior paranasal pillars; Rightmire 1998). And although other differences in cranial form may exist between the two regions, certain characters including angulation of the petrous temporal, size and orientation of the tympanic tend to unite the entire sample, albeit with a range of size variation for some of these features.

Regardless of one’s position on the taxonomic significance of African versus Asian morphology, during this period, the view of H. erectus (or H. ergaster) as relatively large brained and large bodied became standard. Large brain size \( (x = 1,000 \text{ cm}^3) \) continued to be the rubric for H. erectus despite evidence from all parts of its range of individuals with markedly smaller brains, and despite the smaller brain sizes of the African remains. Furthermore, the question of stasis or gradual brain size increase through time became a prevalent question in light of new ideas about punctuated equilibrium (Wolpoff 1984; Rightmire 1985, 1986). Gradual increase in brain size at a rate of about 160 cm\(^3\)/Myr would ultimately be shown to be the case (Leigh 1992; Antón and Swisher 2001). The implication of brain size increase for the costs incurred by both growing infant and mother also became a theme (Martin 1996). The argument seemed to swing in favor of a more humanlike ontogeny than in the australopiths; however, the interpretations ranged between ontogeny being more apelike or, alternatively, more humanlike (Mann 1981; Smith 1993). These life history debates continue into the present period with new data pointing, in some cases, toward a more apelike ontogenetic pattern (Dean et al. 2001) and other data being more supportive of a humanlike brain ontogeny (see Leigh, 2006 but contra Coqueugniot et al. 2004, Hublin and Coqueugniot 2006; see below).

On the basis of KNM-WT 15000 and isolated African postcranial remains, H. erectus came to be viewed as of modern human size and proportions. These body size inferences supported the argument for a more humanlike pattern of growth and have important implications for reconstructions of various life
history parameters that would be more fully explored during the next period of
discovery (Bogin and Smith 1996; Clegg and Aiello 1999; Dean et al. 2001; Aiello
and Key 2002; Aiello and Wells 2002; Antón 2002b; Antón and Leigh 2003).
However, intraspecific and regional variation in body size deserves more atten-
tion than the focus on the fabulously complete KNM-WT 15000 allowed (Antón
2004), an issue that we will discuss later.

Also during this period, the case grew for a relatively late hominin migration
out of Africa, at or after 1 Ma (Pope 1983; Klein 1989). Such a dispersal date
significantly postdated the appearance of an Acheulean toolkit and was argued to
have been facilitated if not induced by changes in hominin technological prowess
(Klein 1989). The argument for a late dispersal placed the earliest Indonesian
hominins at about 1 Ma, juxtaposed in time between the Early Pleistocene
African remains and the Middle Pleistocene Chinese H. erectus. These ages
stood in contrast to the earlier geological opinions based on stratigraphic context
and to some early, if imprecise, K–Ar dates for Java (Jacob and Curtis 1971). This
late dispersal also stood in contrast to the apparently early changes in body size
and shape and distribution of archeological sites, both of which argued for home
range expansion and a foraging shift in early H. erectus (Shipman and Walker
1989).

Based on these finds from the second period of discovery, H. erectus was
firmly placed in the hominin lineage (Figure 11.2). A definition of H. erectus sensu lato continued to emphasize a moderate brain size, although now recogniz-
ing that 850 cm³ might be a minimum size (at 727 cm³ the cranial capacity of OH
12 was estimated well below this, but this fragmentary specimen was rarely
considered in definitions of the taxon), a thick walled, angular vault, and cranial
base characters. The definition began to consider the issue of geographic variation
in cranial characters in the form of the H. erectus/H. ergaster question but tended
not to address temporal variation (except in brain size), regional variation not
related to the Africa/Asia dichotomy, local or idiosyncratic variation. Body size
and proportions were argued to be humanlike for the first time in the hominin
lineage.

In this period, the debate about H. erectus seems very much focused on
the distinction between H. erectus and H. sapiens, whereas the distinction between
H. erectus and H. habilis rested largely on the absolutely larger brain of H. erectus.
Difficulty in finding autapomorphies that united the worldwide sample of
H. erectus resulted in the H. erectus/H. ergaster distinction. However, reexamina-
tion of the perceived apomorphies also led to the conclusion that H. erectus and
H. sapiens could not be distinguished cladistically (Hublin 1986). Despite this, a
grade concept of H. erectus was still considered useful by these and other authors.
In the 1990s, the discovery of small H. erectus crania has focused attention on the
Defining *Homo erectus*: size considered

*H. habilis/H. erectus* distinction, and in this discussion the effects of size and scale become important.

### 11.2.3 The recent period (1990s to present)

The most recent period of discovery is punctuated by two kinds of new data: spectacular new fossils and archeological sites at ages and in places not previously anticipated (Gabunia and Vekua 1995; Gabunia et al. 2000a; Vekua et al. 2002; Zhu et al. 2004) and revisions of the chronostratigraphic framework of known sites that challenge the view of a late dispersal and early demise of *H. erectus* by arguing for a dispersal from Africa as early as 1.8 Ma and persistence in island Southeast Asia until at least 50 ka, if not later (Swisher et al. 1994, 1996; Larick et al. 2001).

We remain unsure whether hominin dispersals from Africa to East Asia began at 1.6 Ma or as early as 1.8 Ma; however, the majority of data point to dispersal substantially before 800 ka, the accepted position just a decade ago. Here we provide the “highlights” of the longer chronology, a detailed discussion of the revision of the chronostratigraphic framework (and its critics) is provided elsewhere (Antón 2002a, 2003; Antón and Swisher 2004). New archeological and paleontological sites suggest that the earliest occupation of Europe occurred by at least 800 ka (Carbonell et al. 1999), of Western Asia by 1.7 Ma (Gabunia et al. 2000a), and of China by 1.6 Ma (Zhu et al. 2004). Radiometric age estimates of the entire stratigraphic section in the Sangiran Dome region of Indonesia, which yielded the bulk of the Indonesian fossils from the previous two periods, suggest the first dispersal into Southeast Asia occurred by 1.6 Ma and entirely preceded 1 Ma (Swisher et al. 1994, 1997; Larick et al. 2001). In eastern Java, at Perning, an age of 1.8 Ma is suggested (Swisher et al. 1994).

The younger end of the age spectrum has also changed with Late Pleistocene radiometric ages of perhaps as little as 25 ka for *H. erectus* on Java (Swisher et al. 1996) and if further investigation supports both the age and taxonomic placement of the recently discovered Liang Bua hominins, of perhaps as little as 18 ka on Flores (Brown et al. 2004). It is worth noting that this revised chronology for Southeast Asia supports the geologically based views of the pre1950s period that argued that the *H. erectus* sites spanned from the earliest to the latest Pleistocene (de Terra 1943; von Koengiswald 1962).

The revised chronology has a direct effect on our understanding of the timing and causes of the first hominin migrations from Africa (Bar-Yosef and Belfer-Cohen 2000; Gabunia et al. 2000a; Antón et al. 2001, 2002; Mithen and Reed 2002; Antón and Swisher 2004). Unlike our understanding just a decade
ago, Early Pleistocene dispersal argues for a more ecomorphological view of this process. It is likely that \textit{H. erectus} dispersed as part of the ecological shifts that also drove the dispersal of other mammals of the same age. While it is likewise probable that \textit{H. erectus} had undergone a shift in foraging strategy, as witnessed by its larger body size and presumably home range (Shipman and Walker 1989; Leonard and Robertson 1997, 2000), and that technology played a role in this shift, it is unlikely that technological innovation was the sole impetus for \textit{H. erectus} dispersal capability (Antón et al. 2002).

Beyond the timing of the first dispersal, however, the new chronology poses several issues for the interpretation of the fossil sample of \textit{H. erectus}. The earlier view of a sequential time series from Africa to Southeast Asia to continental Asia provided the possibility of considering the regional characters as a series of evolutionary modifications on a theme and made comparison in similar time frames impossible. Now, the vast time span in Southeast Asia argues that, at least initially, a conservative approach to comparisons across geographic regions would be to analyze samples that are broadly contemporaneous. In some cases this is possible, as in the evaluation of the very earliest \textit{H. erectus} in Africa, Western Asia, and Southeast Asia, albeit with limited sample sizes (Antón and Indriati 2002). And it is likewise possible to evaluate trends through time in single regions (Antón 2002a; Kaifu et al. 2005). However, between some regions, such as China and Indonesia, there are no comparably aged specimens and often the best preserved fossils from each region are not broadly contemporaneous, thus requiring a certain amount of interpolation of expectations (Figure 11.4). Additionally, for some samples, particularly the postcranial sample, the number of specimens is simply not sufficient to consider anything but the entire, time-transgressive, sample.

Beginning in the 1990s, a series of fossil finds have brought into focus the assumptions that we had made previously about the large size of \textit{H. erectus}, both cranially and postcranially. The most spectacular of the hominin finds have been the multiple individuals from Dmanisi in the Republic of Georgia. All the Dmanisi crania are said to express anatomical characters of \textit{H. erectus} but with small cranial capacities ranging in adults from 650 to 780 cm$^3$. Likewise, the recently discovered Ilfer (KNM-ER 42700; Leakey et al. 2003) and Olorgesailie (KNM-OL 45500; Potts et al. 2004) hominins from Kenya and perhaps the Liang Bua specimen from Flores (Brown et al. 2004) are all of smaller size than the more complete of the \textit{H. erectus} specimens might otherwise lead one to expect. The calvaria from Ilferet also bears directly on the African versus Asian controversy as it exhibits several “Asian” characters (Leakey et al. 2003; Spoor et al. 2005). Importantly, in the case of Dmanisi, there are also postcranial remains that provide some estimate of body size and a means for scaling body and brain size.
As a group these hominins, together with the neglected OH 12, argue that size and scaling are issues of some importance in understanding the \textit{H. erectus} radiation and the definition of the taxon (Antón 2003, 2004; Spoor et al. 2005). If we now consider a definition in light of variation in body and brain size, it becomes clear that there is at least the possibility that some features of that definition may scale with size (either body or brain). Given the nonmetric characters of the Dmanisi specimens (and OH 12), it is necessary to dismiss the idea that all adult \textit{H. erectus} had brain sizes in excess of 800 cm$^3$. Indeed, an absolute rubric for brain size without some other scaling factor relative to body size seems unwarranted. We also might intuitively anticipate that features such as cranial thickness or tympanic thickness may scale with size. But other structural
relationships, such as the position of the glenoid fossa relative to the brain and even the angulation of the vault, may also reflect scaling relationships (Spoor et al. 2005). The extent to which size varies among the regional samples of *H. erectus sensu lato* and how it affects the expression of characters is reviewed below.

### 11.3 What has size got to do with it?

Given the known sample of *H. erectus sensu lato*, the extent to which size influences the expression of characters that have in the past been included in definitions of the taxon becomes relevant. In addition, the extent to which size varies temporally and geographically and how this may influence our interpretations of the sample are also issues worth pursuing. In this section, we consider both brain size and body size scaling.

#### 11.3.1 Body size variation among *H. erectus*

The difficulties with reconstructing body size in the fossil record are well known; there is relatively little postcranial material, and less yet that is reliably associated with cranial remains identified to species. *H. erectus* represents a somewhat extreme case in this regard. Associated cranial and postcranial remains from the same individual amount to three individuals from the early African record (KNM-WT 15000, KNM-ER 1808, and KNM-ER 803), additional “large” postcranials of the right age from Koobi Fora are assigned somewhat by default to *H. erectus* (e.g., KNM-ER 737). Other remains of the right age from Olduvai Gorge may be similarly assigned (e.g., OH 28, OH 34). Remains in Asian localities with only a single hominin species can also be referred to *H. erectus* (e.g., from Dmanisi, Zhoukoudian, Ngandong, Sangiran; see Chapter by Tattersall for a dissenting viewpoint).

For the investigation of life history variables, energetic requirements, and so on, it is useful to attempt to estimate absolute body size and proportions (McHenry and Coffing 2000; Ruff 2002). Given the small number of associated partial skeletons, we do not consider body proportions here. Two means of assessing absolute body size, stature and weight, are both important variables for assessing the taxon. However, body mass estimates require more information regarding body shape than do stature estimates (which also require several assumptions; Ruff et al. 1997; Ruff 2002). In the best case, both leg length and bi-iliac breadth can be used to estimate body mass. Femoral head size has also proven useful as an estimator of body mass. The Asian *H. erectus* fossil record
lacks these elements and we therefore limit our discussion to stature estimates here, except as is necessary to compare with previous studies.

Absolute stature can be estimated from the skeleton in a variety of ways. The most useful means for isolated skeletal elements is the use of regression analyses that predict stature from single elements of the lower limb (Pearson 1898; Telkkä 1950; Trotter and Gleser 1952, 1958, 1977). The accuracy of the estimate is dependent on the similarity of proportions between the known sample and the individual whose stature is to be estimated, this remains a significant source of error for hominin samples in which proportions are not known. As with many fossil assemblages, most of the *H. erectus* postcranial remains are not only isolated but also fragmentary; in particular in the case of the Asian record of the species. Visual reconstruction and substitution methods are often used to predict total long bone length in fossils, although the range of error is difficult to predict (McHenry 1991). Femoral length can be predicted with some accuracy in humans from femoral head size (McHenry 1991), although the technique assumes a similar relationship between joint size and limb length throughout the *Homo* lineage. An alternate method developed for forensic anthropological use is the creation of segment based regression analyses that use portions of the long bone to predict its absolute length (Steele and McKern 1969; Steele 1970; Jacobs 1992; Wright and Vásquez 2003M; Fellmann 2004). We report some of these estimates for Asian *H. erectus* femora base on Asian human femoral segment data (after Fellmann 2004) and their correlated stature predictions using the femur/stature ratio equation \[ \text{stature (cm)} = \frac{\text{femur length (cm)}}{C} \times 100/26.74 \] of Fedelsman et al. (1990) and stature estimates based on Asian human regression equations. We also report estimates of *H. erectus* stature based on other, previously published techniques. We report estimates of *H. erectus* stature based on other, previously published techniques (Table 11.2).

Absolute stature estimates based on the tibia and femur vary by region among *H. erectus* (Table 11.3). As might be predicted on the basis of ecological laws for heat retention and dissipation, tropical *H. erectus* from Africa and Indonesia have longer lower limb lengths and related statures, whereas more northerly located *H. erectus*, including those from Georgia and China, are substantially shorter. The ranges do not overlap between tropical and temperate *H. erectus* samples. However, sample sizes are quite small. The African sample has the largest mean size and largest individual stature value. Its range nearly encompasses that of the Indonesian sample. There is a suggestion that later Indonesian *H. erectus* (the Ngandong tibia) differs more in body size from African *H. erectus* (the Nariokotome tibia) than the earlier Indonesian sample (Trinil femora) differ from African *H. erectus* (Koobi Fora and Olduvai femora). If true, this would suggest a decrease in body size in Indonesian *H. erectus* with time. However, with very small sample sizes we cannot make a great deal of this apparent difference.
<table>
<thead>
<tr>
<th>Specimen</th>
<th>Taxon</th>
<th>Length (mm)</th>
<th>Midshaft AP (mm)</th>
<th>Midshaft ML (mm)</th>
<th>Stature (cm)</th>
<th>Stature by Fem/Stature ratio (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Femora</strong></td>
<td>Bicondylar</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KNM-ER 1472</td>
<td>H. sp.</td>
<td>400</td>
<td>25.1</td>
<td>26.4</td>
<td>149</td>
<td>150</td>
</tr>
<tr>
<td>KNM-ER 1481</td>
<td>H. sp.</td>
<td>395</td>
<td>22.4</td>
<td>25.6</td>
<td>147</td>
<td>148</td>
</tr>
<tr>
<td>KNM-ER 3728</td>
<td>H. sp.</td>
<td>390</td>
<td>20.1</td>
<td>24.7</td>
<td>145</td>
<td>146</td>
</tr>
<tr>
<td>KNM-ER 736</td>
<td>H.e.</td>
<td>500</td>
<td>36.1</td>
<td>37.7</td>
<td>180</td>
<td>186</td>
</tr>
<tr>
<td>KNM-ER 737</td>
<td>H.e.</td>
<td>440</td>
<td>27.1</td>
<td>32.4</td>
<td>160</td>
<td>164</td>
</tr>
<tr>
<td>KNM-ER 1808</td>
<td>H.e.</td>
<td>480</td>
<td>–</td>
<td>–</td>
<td>173</td>
<td>180</td>
</tr>
<tr>
<td>KNM-WT 15000</td>
<td>H.e.</td>
<td>429 at death</td>
<td>24.5</td>
<td>24.3</td>
<td>159 at death</td>
<td>160</td>
</tr>
<tr>
<td>KNM-WT 15000</td>
<td>H.e.</td>
<td>517 adult</td>
<td>–</td>
<td>–</td>
<td>185 adult</td>
<td>193</td>
</tr>
<tr>
<td>OH 28</td>
<td>H.e.</td>
<td>448</td>
<td>24.7</td>
<td>32.7</td>
<td>167</td>
<td>168</td>
</tr>
<tr>
<td>OH 34</td>
<td>?H.e.</td>
<td>430</td>
<td>–</td>
<td>–</td>
<td>158</td>
<td>161</td>
</tr>
<tr>
<td>Trinil I</td>
<td>?H.e.</td>
<td>455</td>
<td>29</td>
<td>28</td>
<td>163</td>
<td>170</td>
</tr>
<tr>
<td>Trinil II</td>
<td>?H.e.</td>
<td>447</td>
<td>26.2</td>
<td>27.1</td>
<td>167</td>
<td>167</td>
</tr>
<tr>
<td>Trinil III</td>
<td>?H.e.</td>
<td>433</td>
<td>25.4</td>
<td>27.7</td>
<td>162</td>
<td>162</td>
</tr>
<tr>
<td>Zhouk I</td>
<td>H.e.</td>
<td>378</td>
<td>27.1</td>
<td>29.7</td>
<td>150*</td>
<td>141</td>
</tr>
<tr>
<td>Zhouk IV</td>
<td>H.e.</td>
<td>413</td>
<td>25</td>
<td>29.3</td>
<td>150*</td>
<td>154</td>
</tr>
<tr>
<td><strong>Tibia</strong></td>
<td>Maximum</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KNM-WT 15000</td>
<td>H.e.</td>
<td>380 at death</td>
<td>24.0</td>
<td>20.4</td>
<td>160</td>
<td>–</td>
</tr>
<tr>
<td>KNM-WT 15000</td>
<td>H.e.</td>
<td>475 as adult</td>
<td>–</td>
<td>–</td>
<td>184</td>
<td>–</td>
</tr>
<tr>
<td>Ngandong A</td>
<td>H.e.</td>
<td>–</td>
<td>37.1</td>
<td>27.0</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Ngandong B</td>
<td>H.e.</td>
<td>(360)</td>
<td>29.5</td>
<td>21.2</td>
<td>158</td>
<td>–</td>
</tr>
<tr>
<td><strong>Third Metatarsal</strong></td>
<td>Maximum</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dmanisi D2021</td>
<td>H.e.</td>
<td>60</td>
<td>–</td>
<td>–</td>
<td>148</td>
<td>–</td>
</tr>
</tbody>
</table>

Data are from: Africa—McHenry (personal communication, 1992, 1994); McHenry and Coffing (2000); Walker and Leakey (1993); Fellmann (2004); China—Weidenreich (1941); Wu and Poirier (1995); Fellmann (2004); Indonesia—Santa Luca (1980); Fellmann (2004); Georgia—Gabunia et al. (2000b).

*Weidenreich, using Manouvrier's and Pearson's regressions, published an estimated stature of 156 cm (1941 p 34).
<table>
<thead>
<tr>
<th></th>
<th>Africa</th>
<th>Georgia</th>
<th>China</th>
<th>Indonesia</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>H. erectus s. l.</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Femur length (mm)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>459</td>
<td>–</td>
<td>395</td>
<td>445</td>
</tr>
<tr>
<td>Range (n)</td>
<td>430–500 (5)*</td>
<td>–</td>
<td>378–413 (2)</td>
<td>433–455 (3)</td>
</tr>
<tr>
<td>Tibia length (mm)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>475</td>
<td>–</td>
<td>–</td>
<td>360</td>
</tr>
<tr>
<td>Range (n)</td>
<td>475</td>
<td>–</td>
<td>–</td>
<td>360</td>
</tr>
<tr>
<td>Stature (cm)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>174</td>
<td>148</td>
<td>148</td>
<td>164</td>
</tr>
<tr>
<td>Range (n)</td>
<td>161–186 (6)**</td>
<td>148</td>
<td>141–154 (2)</td>
<td>158–170 (4)</td>
</tr>
<tr>
<td><strong>H. sapiens</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Femur length (mm)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>African American¹</td>
<td>European American¹</td>
<td>China⁴</td>
<td>Andamanese⁷</td>
</tr>
<tr>
<td></td>
<td>474 male</td>
<td>457 male</td>
<td>439 male</td>
<td>385 mixed sex</td>
</tr>
<tr>
<td></td>
<td>437 female</td>
<td>430 female</td>
<td>Native American⁵</td>
<td></td>
</tr>
<tr>
<td>Range (n)</td>
<td>Native American⁵</td>
<td>440 male</td>
<td>413 female</td>
<td>374–479</td>
</tr>
<tr>
<td>Tibia length (mm)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>African American¹</td>
<td>China⁴</td>
<td>Andamanese⁷</td>
<td></td>
</tr>
<tr>
<td></td>
<td>387 male</td>
<td>362 male</td>
<td>325 mixed sex</td>
<td></td>
</tr>
<tr>
<td></td>
<td>354 female</td>
<td>Native American⁵</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>371 male</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>346 female</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Range (n)</td>
<td>Native American&lt;sup&gt;5&lt;/sup&gt;</td>
<td>298–401</td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>----------------------------</td>
<td>---------</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Stature (cm)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td>African American&lt;sup&gt;1&lt;/sup&gt;</td>
<td>European American&lt;sup&gt;1&lt;/sup&gt;</td>
<td>China&lt;sup&gt;4&lt;/sup&gt;</td>
<td>Phillipines&lt;sup&gt;8&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>172 male</td>
<td>170 male</td>
<td>169 male</td>
<td>160 male</td>
</tr>
<tr>
<td></td>
<td>161 female</td>
<td>160 female</td>
<td>Evenki&lt;sup&gt;6&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Turkana&lt;sup&gt;2&lt;/sup&gt;</td>
<td>Netherlands&lt;sup&gt;2&lt;/sup&gt;</td>
<td>161 male</td>
<td></td>
</tr>
<tr>
<td></td>
<td>174 male</td>
<td>182 male</td>
<td>149 female</td>
<td></td>
</tr>
<tr>
<td></td>
<td>162 female</td>
<td>168 female</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mbuti&lt;sup&gt;2&lt;/sup&gt;</td>
<td>Portuguese&lt;sup&gt;3&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>145 male</td>
<td>168 male</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>138 female</td>
<td>158 female</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Range (n)</strong></td>
<td></td>
<td>Portuguese&lt;sup&gt;3&lt;/sup&gt;</td>
<td>Evenki&lt;sup&gt;6&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>153–185 male</td>
<td>130–175 male</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>145–175 female</td>
<td>140–161 female</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>1</sup>African American and European American mean values for the Terry collection following Trotter and Gleser.<br>
<sup>2</sup>Turkana and Dutch mean values from Bogin (1991 p 106 after Little 1983 and Mann 1962).<br>
<sup>3</sup>Portuguese values from De Mendonca (2000).<br>
<sup>4</sup>Mean values from Krogman (1978 p 176).<br>
<sup>5</sup>Date from Berschini and Haversat (1980) for precontact Central Californians.<br>
<sup>6</sup>Data Courtesy of Bill Leonard.<br>
<sup>7</sup>Data from Stock and Pfeiffer (2001).<br>
<sup>8</sup>Data from Murray (2002).
The entire *H. erectus* sample is within the size range of living human stature as represented by the extremes of the Mbuti pygmies from Africa and the Dutch (*Table 11.3*). The *H. erectus* patterns between regions that we interpret as evidence of clinal variation in *H. erectus* body size compare favorably with the pattern of variation in modern *H. sapiens*. The regional samples are also broadly similar to regional samples of recent modern humans, although the Dmanisi sample is on the small end of the range (*Table 11.3*). Likewise, Ruff (2002) has argued that body shape (bi-iliac breadth to ulna length) in Pleistocene *Homo* also follows clinal trends similar to those seen in modern humans. Ruff’s (2002) data suggest that high latitude Pleistocene *Homo* is 10% larger than are high latitude recent humans, a distinction that we do not see between *H. erectus* in Georgia and China and recent humans from temperate climates.

Before turning to the evaluation of brain size variation in *H. erectus*, it is important to evaluate whether there is a trend for body size increase (or decrease) through time in *H. erectus*. As will be clear from *Figure 11.5a*, although there is a trend toward stature increase with time across hominins, stature does not increase with time within *H. erectus*. The flat slope is due mostly to the fact that the largest statures, those already within the modern human range, are present in the earliest Pleistocene sample of *H. erectus*. We should note that this applies to stature only and that body mass through time remains an unknowable variable.

---

**Figure 11.5**

(a) Relationship between Stature and geological age (in millions of years) in hominins.  
(b) Relationship between body mass and geological age (in millions of years) in *H. erectus* as estimated by Ruff et al. (1997)
for *H. erectus* due to the absence of either femoral head remains or bi-iliac breadths for Asian *H. erectus*.

Ruff et al. (1997) noted a similar body *mass* increase across the genus *Homo*, including archaic and modern *H. sapiens*, up to about 50 ka. By implication, this would include increase in size within *H. erectus* as well, although their analyses do not directly speak to this point as they combined specimens across all *Homo* species by time interval, averaging body *size* and cranial capacity within interval. There are several possibilities for the apparent difference between our findings. There may be both a body mass increase but no change in stature in *H. erectus* through time, suggesting that the two are tracking somewhat different things. Alternatively, within *H. erectus* neither body mass nor stature may increase with time suggesting that the genus wide trend for mass increase found by Ruff and colleagues is driven by the body mass increase in archaic *H. sapiens* skeletons and/or by differences in included specimens, samples, and age data between our studies. We tend to favor the latter option given our expectation that the proportions of both the early and late Asian *H. erectus* from Trinil and Ngandong, Java respectively are liable to mirror those of the tropical body proportions of the early African remains, and given similar body proportions their statures should translate to similar body masses. Thus, the latest surviving *H. erectus*, with relatively tall statures, should be of similar or slightly smaller mass as the earliest African and Asian *H. erectus*. Given that these late samples also have the largest brain sizes, similar body sizes would support the idea of gradual brain size increase in both absolute and relative terms. We evaluate this possibility using the estimated body mass data for *H. erectus* only from Ruff et al.’s (1997, supplementary data) analyses. Even without adjusting their geological age data to reflect more recent revisions of the chronostratigraphic framework, there is no correlation between body mass and geological age within *H. erectus* and if anything there is a decrease in body mass with time (Figure 11.5b). When geological age data are adjusted to reflect current knowledge, there is still no evidence for body mass change with time. When newly discovered *H. erectus* and estimates for additional known *H. erectus* are added to the sample of Ruff et al. (1997), there is likewise no evidence for body mass increase with time.

### 11.3.2 Brain size through time and space

As mentioned, the absolute range of variation in brain size in *H. erectus* and the geochronological scale of the species have both changed in the past 15 years. Although it is widely recognized that brain size has increased through the evolutionary history of our genus, whether *H. erectus* exhibited a rapid brain
increase followed by relative stasis (Eldredge 1985; Rightmire 1985, 1986) or a gradual increase in brain size through time (Leigh 1992; Antón and Swisher 2001) has been contested. The correlation, if any, between the range of individual brain sizes and time is also a point of debate.

We have previously established that there is a directional trend in absolute brain size increase in Asian $H$. erectus (Antón and Swisher 2001). Based on a revised geochronological framework and established cranial capacities for 22 fossil hominins (Table 11.4), Kendall’s Tau (a nonparametric test of randomness in rank-ordered data; Konigsberg 1990) suggests that there is a directional change (increase) in cranial capacity with decreasing geological age [$\text{Tau} = 0.509; p < 0.01$; based on equations of Sokal and Rohlf (1995 pp 594–597) and probability values from Table S of Rohlf and Sokal (1995 p 127)]. An ordinary least squares (LS) regression of individual data points and average cranial capacities at different time intervals yields slopes significantly different from zero (Table 11.5), and these results are not significantly affected when the geologically oldest (Mojokerto projected adult value) and youngest (Ngandong) data points are removed. Based on the revised age-calibration of the fossils, brain size increased in Asian $H$. erectus at the average rate of about 160 cm$^3$/Myr (Antón and Swisher 2001). Sample sizes preclude establishing rates of increase for Africa or comparing between African and Asian rates. However, the cranial capacities and ages of the African specimens are similar to those from Asian and unlikely to alter rates much.

Differences between $H$. erectus and archaic $H$. sapiens rates of brain size increase reflect an exponential increase in brain size at the origin of $H$. sapiens (Henneberg 1987, 2001) and suggest the possibility of a different ontogenetic trajectory for brain growth in the two species. Leigh (1992), using a worldwide sample of early $H$. sapiens, showed a rate of increase of about 860 cm$^3$/Myr, although this rate may be somewhat mitigated by concomitant body mass increase in archaic $H$. sapiens (Ruff et al. 1997). Leigh’s rate for archaic $H$. sapiens is over four times the rate for $H$. erectus established by us ($\sim$135–177 cm$^3$/Myr). Our rates of increase suggest an even greater disparity between rates of brain size increase in Asian $H$. erectus and in early $H$. sapiens than found by Leigh (1992). Although these differences suggest that ontogenetic brain trajectories may differ between the species, it is not clear at what point during brain growth that trajectory might change. If the new developmental age estimates for Mojokerto are correct (Coqueugniot et al. 2004), then brain size in young $H$. erectus ($\sim$660 cm$^3$/690 g) is within the modern human range for its age (human range 600–1200 g at 1–1.5 years), suggesting that any change in growth trajectory may have occurred as a result of a slowdown, or cessation in brain growth during the juvenile or
adolescent period of *H. erectus* rather than in the pre- or just postnatal period (see also Leigh 2006). This would suggest that *H. erectus* was likely to be secondarily altricial and would also support an earlier age at maturation in *H. erectus* than *H. sapiens* (Smith 1993; Dean et al. 2002).

<table>
<thead>
<tr>
<th>Specimen</th>
<th>Cranial capacity</th>
<th>Source of capacity</th>
<th>Age (range)</th>
<th>Source of age (method)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Indonesia</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mojokerto adult</td>
<td>800</td>
<td>Antón (1997)</td>
<td>1.81 Ma</td>
<td>Swisher et al. (1994) (40Ar/39Ar)</td>
</tr>
<tr>
<td>Sangiran-4</td>
<td>908</td>
<td>Holloway (1981)</td>
<td>~1.6 Ma*</td>
<td>As above</td>
</tr>
<tr>
<td>Sangiran-2</td>
<td>813</td>
<td>As above</td>
<td>~1.5 Ma*</td>
<td>Swisher unpublished</td>
</tr>
<tr>
<td>Sangiran-17</td>
<td>1004</td>
<td>As above</td>
<td>1.3 Ma*</td>
<td>As above</td>
</tr>
<tr>
<td>Sangiran-10</td>
<td>855</td>
<td>As above</td>
<td>1.2 Ma*</td>
<td>As above</td>
</tr>
<tr>
<td>Sangiran-12</td>
<td>1059</td>
<td>As above</td>
<td>1.1 Ma*</td>
<td>As above</td>
</tr>
<tr>
<td>Trinil</td>
<td>940</td>
<td>As above</td>
<td>&lt;1 Ma</td>
<td>As above</td>
</tr>
<tr>
<td>Ngandong1</td>
<td>1172</td>
<td>Holloway (1980)</td>
<td>0.05 Ma(0.027–0.053 Ma)</td>
<td>Swisher et al. (1996) (U-series/ESR)</td>
</tr>
<tr>
<td>Ngandong-6</td>
<td>1251</td>
<td>As above</td>
<td>As above</td>
<td>As above</td>
</tr>
<tr>
<td>Ngandong-7</td>
<td>1013</td>
<td>As above</td>
<td>As above</td>
<td>As above</td>
</tr>
<tr>
<td>Ngandong-10</td>
<td>1135</td>
<td>As above</td>
<td>As above</td>
<td>As above</td>
</tr>
<tr>
<td>Ngandong-11</td>
<td>1231</td>
<td>As above</td>
<td>As above</td>
<td>As above</td>
</tr>
<tr>
<td>Ngandong-12</td>
<td>1090</td>
<td>As above</td>
<td>As above</td>
<td>As above</td>
</tr>
<tr>
<td>Sambungmacan-</td>
<td>1035</td>
<td>Pope (1988)</td>
<td>As above</td>
<td>As above</td>
</tr>
<tr>
<td><em>China</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hexian</td>
<td>1025</td>
<td>Rightmire (1985)</td>
<td>0.20 Ma (0.15–0.27 Ma)</td>
<td>Chen et al. (1987) (U-series)</td>
</tr>
<tr>
<td>Zhoukoudian</td>
<td>1030</td>
<td>Weidenreich (1943)</td>
<td>0.42 Ma</td>
<td>Huang et al. (1993) (ESR)</td>
</tr>
<tr>
<td>Skull II</td>
<td>915</td>
<td>As above</td>
<td>0.58 Ma</td>
<td>As above</td>
</tr>
<tr>
<td>Skull III</td>
<td>1140</td>
<td>Chiu et al. (1973)</td>
<td>0.30 Ma (0.25–0.28 and 0.31–0.38 Ma)</td>
<td>Grün et al. (1997, ESR) Huang et al. (1993, ESR, U-series) Zhao et al. (1985, U-series)</td>
</tr>
<tr>
<td>Skull VI</td>
<td>850</td>
<td>Weidenreich (1943)</td>
<td>0.42 Ma</td>
<td>Huang et al. (1993) (ESR)</td>
</tr>
<tr>
<td>Skull X</td>
<td>1225</td>
<td>As above</td>
<td>As above</td>
<td>As above</td>
</tr>
<tr>
<td>Skull XI</td>
<td>1015</td>
<td>As above</td>
<td>As above</td>
<td>As above</td>
</tr>
<tr>
<td>Skull XII</td>
<td>1030</td>
<td>As above</td>
<td>As above</td>
<td>As above</td>
</tr>
</tbody>
</table>

*Estimated age based on bracketing tuffs of the Bapang (Kabuh) formation of Sangiran between 1.07 and 1.58 Ma. See text Antón and Swisher (2001) for detailed discussion.
The finds of the 1990s/2000s show that brain size in *H. erectus sensu lato* ranges from about 650 to 1250 cm³. As the trend toward increasing brain size through time suggests, there is a slight shift in this total range through time. The Early Pleistocene sample (from /C24 1.0 to 1.8 Ma) ranges from 650 to 1067 cm³, the Middle Pleistocene sample (from China) ranges from about 850 to 1225 cm³, and the Late Pleistocene sample (from Ngandong and Sambungmacan) ranges from about 915 to 1251 cm³.

### 11.3.3 Brains relative to body size

Having established that there is no temporal trend in stature through time within *H. erectus* but that there is a small increase in brain size through time, the scaling

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R²</td>
<td>Slope</td>
</tr>
<tr>
<td>Early <em>H. sapiens</em></td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Combined African/Asian <em>H. erectus</em></td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Ngandong alone</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Asian <em>H. erectus</em> w/o Ngandong</td>
<td>0.36</td>
<td>–</td>
</tr>
<tr>
<td>Asian <em>H. erectus</em> w/ Ngandong</td>
<td>0.49</td>
<td>–</td>
</tr>
<tr>
<td>Asian <em>H. erectus</em> w/ Ngandong and w/o Mojokerto</td>
<td>0.41</td>
<td>–</td>
</tr>
<tr>
<td>Indo <em>H. erectus</em> w/o Ngandong</td>
<td>0.28</td>
<td>–</td>
</tr>
<tr>
<td>Indo <em>H. erectus</em> w/ Ngandong</td>
<td>0.67</td>
<td>–</td>
</tr>
<tr>
<td>China <em>H. erectus</em></td>
<td>0.12</td>
<td>–</td>
</tr>
<tr>
<td>Asian <em>H. erectus</em> mean values w/Ngandong</td>
<td>0.76</td>
<td>–</td>
</tr>
</tbody>
</table>

Differences between studies are due to revisions in the geochronology since the Leigh (1992) study.

Defining *Homo erectus*：size considered
relationship between body and brain size provides another clue as to the position of *H. erectus* relative to *H. sapiens* and earlier taxa. Again, this exercise is hampered by the small sample sizes but can be approached in a couple of ways. First, we can look at the scaling relationship between the size of a particular element, usually the femur, and brain size. This, of course, limits us to individuals that retain both these elements or to chimeras of femora and crania from specific localities (say Olduvai Gorge). Alternatively, we can from various elements estimate a probable stature and scale this to brain size. This has the advantage of increasing our sample size somewhat since we could include estimates from individuals with either a tibia or a femur (or other element), for example. In this case in particular, it allows us to include the Dmanisi fossil sample. However, it has the disadvantage that the stature estimates themselves introduce another layer of error into the analyses (as they rely on scaling relationships in humans to predict stature from long bones, and these relationships vary by body proportion, which is unknown for some geographic samples of *H. erectus* and because the range of error is compounded by each additional estimate [see Smith 1996]).

The samples used here have been constructed as per Table 11.6, by using associated individuals and chimeras of brain and body size from particular sites. Where multiple crania exist at a site, the largest and smallest capacities are both matched with the postcranial estimate. When ranges of both exist, crania and postcrania are matched according to size and, to the extent possible, geological age. Changes to these chimeras may affect results but are unlikely to change ranges. Despite the potential issues with this system, we prefer this method over the use of mean species values for stature and cranial capacity as we are concerned primarily with issues of size change within *H. erectus* which requires individual data points. Figure 11.6 illustrates the relationship between cranial size and femur length for the specimens indicated in Table 11.6, whereas Figure 11.7 illustrates the relationship between cranial size and stature. Modern human femoral length data represent a worldwide sample (from Brown 2005; Berschini and Haversat 1980), whereas stature data are calculated from data for precontact Central California Native American samples (Berschini and Haversat 1980) using regression formula for Asian modern humans (Trotter 1970).

Although the fossil sample sizes are quite small, *H. erectus* individuals appear to fall on the same scaling relationship between femur length and brain size as all other hominins except ourselves (Figure 11.6; Table 11.7), although the intraspecific regression for *H. erectus* is not significant, probably due to small sample size. Due to relatively larger brain sizes, the human regression falls above the other hominins, although some *H. erectus* chimeras fall within the *H. sapiens* data cloud. Thus, without knowledge of the position of earlier hominins, distinguishing some *H. erectus* from *H. sapiens* is difficult and no
doubt has led to the claims for *H. erectus* being more “humanlike” in its body size. Alternatively, at the lower end of the range, distinguishing smaller *H. erectus* from larger *H. habilis* is similarly difficult. Note also that the Liang Bua 1 specimen from Flores falls at essentially the same position as the *A. afarensis* datapoint, which in this study is the chimera of the smallest body and brain sizes rather than the species means used by Brown et al. (2004).

Additional *H. erectus* data points can be added if we consider the relationship between stature and brain size (Figure 11.7). And again we see that *H. erectus*

---

**Table 11.6**

<table>
<thead>
<tr>
<th>Specimen for cranial estimate</th>
<th>Capacity</th>
<th>Femur length</th>
<th>Stature estimate</th>
<th>Specimen for postcranial estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>A. afarensis</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AL 162-28</td>
<td>380</td>
<td>3.1/3.1</td>
<td>280</td>
<td>106 AL 288-1</td>
</tr>
<tr>
<td><em>H. habilis</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KNM-ER 1813</td>
<td>510</td>
<td>1.9/1.9</td>
<td>350*</td>
<td>131 KNM-ER 1503**</td>
</tr>
<tr>
<td>KNM-ER 1470</td>
<td>740</td>
<td>1.9/1.9</td>
<td>400</td>
<td>150 KNM-ER 1472</td>
</tr>
<tr>
<td><em>H. erectus</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average of KNM-ER 3733, 3883, WT 15000</td>
<td>853</td>
<td>1.7/1.7</td>
<td>480</td>
<td>173 KNM-ER 1808</td>
</tr>
<tr>
<td>KNM-WT 15000</td>
<td>909</td>
<td>1.5/1.5</td>
<td>429 at death</td>
<td>160 KNM-WT 15000</td>
</tr>
<tr>
<td>OH 9</td>
<td>1067</td>
<td>1.5/1.0</td>
<td>448</td>
<td>167 OH 28</td>
</tr>
<tr>
<td>OH 12</td>
<td>727</td>
<td>1.5/1.2</td>
<td>430</td>
<td>158 OH 34</td>
</tr>
<tr>
<td>Trinil 2</td>
<td>940</td>
<td>0.9/0.9</td>
<td>447</td>
<td>167 Trinil II</td>
</tr>
<tr>
<td>Trinil 2</td>
<td>940</td>
<td>0.9/0.9</td>
<td>433</td>
<td>162 Trinil III</td>
</tr>
<tr>
<td>Zhoukoudian Skull III</td>
<td>915</td>
<td>0.58/0.42</td>
<td>413</td>
<td>158 ZKD Femur IV</td>
</tr>
<tr>
<td>Zhoukoudian Skull VI</td>
<td>855</td>
<td>0.42/0.42</td>
<td>413</td>
<td>141 ZKD Femur I</td>
</tr>
<tr>
<td>Zhoukoudian Skull XII</td>
<td>1225</td>
<td>0.42/0.42</td>
<td>413</td>
<td>158 ZKD Femur IV</td>
</tr>
<tr>
<td>Ngandong 7</td>
<td>1013</td>
<td>0.05/0.05</td>
<td>–</td>
<td>158 Ngandong</td>
</tr>
<tr>
<td>Ngandong 6</td>
<td>1251</td>
<td>0.05/0.05</td>
<td>–</td>
<td>158 Ngandong Tibia B</td>
</tr>
<tr>
<td>Dmanisi 2280</td>
<td>780</td>
<td>1.7/1.7</td>
<td>–</td>
<td>146 Dmanisi 2021</td>
</tr>
<tr>
<td>Dmanisi 2282</td>
<td>650</td>
<td>1.7/1.7</td>
<td>–</td>
<td>146 Dmanisi 2021</td>
</tr>
<tr>
<td><em>H. floresiensis?</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liang Bua 1</td>
<td>417</td>
<td>0.018/0.018</td>
<td>280</td>
<td>106 Liang Bua 1</td>
</tr>
</tbody>
</table>

*Length estimated from femoral head size (McHenry 1991).**Specimen assigned to *H. habilis* or *A. boisei* as are most specimens of this age at Koobi Fora (McHenry 1991), and is near the mean of these specimens at Koobi Fora.
Figure 11.6
Relationship between femur length and cranial capacity in hominins. Filled-circles are *H. erectus*, open-circles are two *H. habilis* and one *A. afarensis* (furthest left below cross), open-triangles are *H. sapiens*, and + is Liang Bua 1

Figure 11.7
Relationship between stature and cranial capacity in hominins. Legend as per Figure 11.6
Table 11.7
Regression results for the relationship between body size and cranial capacity

<table>
<thead>
<tr>
<th></th>
<th>Femur Length</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Stature</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R(^2)</td>
<td>Slope OLS</td>
<td>Intercept</td>
<td>P</td>
<td>n</td>
<td>R(^2)</td>
<td>Slope OLS</td>
<td>Intercept</td>
<td>P</td>
<td>n</td>
</tr>
<tr>
<td><em>H. sapiens only</em></td>
<td>0.14</td>
<td>0.63</td>
<td>3.3</td>
<td>&lt;0.00001</td>
<td>198</td>
<td>0.52</td>
<td>1.5</td>
<td>-0.47</td>
<td>&lt;0.00001</td>
<td>55</td>
</tr>
<tr>
<td><em>H. erectus only</em></td>
<td>0.03</td>
<td>0.42</td>
<td>4.3</td>
<td>ns</td>
<td>8</td>
<td>0.08</td>
<td>0.89</td>
<td>2.3</td>
<td>ns</td>
<td>13</td>
</tr>
<tr>
<td><em>H. erectus and H. habilis</em></td>
<td>0.64</td>
<td>1.36</td>
<td>-1.4</td>
<td>0.006</td>
<td>10</td>
<td>0.47</td>
<td>1.35</td>
<td>-0.03</td>
<td>0.005</td>
<td>15</td>
</tr>
<tr>
<td>All fossil hominins</td>
<td>0.78</td>
<td>1.72</td>
<td>-3.6</td>
<td>&lt;0.00001</td>
<td>11</td>
<td>0.67</td>
<td>1.7</td>
<td>-1.8</td>
<td>&lt;0.00001</td>
<td>16</td>
</tr>
</tbody>
</table>
individuals fall along a scaling relationship with other earlier hominins and the intraspecific regression within *H. erectus* remains insignificant. Nevertheless, the plot suggests that the intraspecific regression for modern humans is parallel with the relationship in *H. erectus*, and perhaps the interspecific hominin regression as well. This would imply that modern humans are characterized by a pure grade shift, unlike the scaling against femoral length (Figure 11.6), where this is not evident. Also, there is no particular distinction in this relationship between the smaller *H. erectus*, those from Dmanisi and OH 12, and the large *H. habilis*, KNM-ER 1470. However, note with caution that in these analyses KNM-ER 1470 is paired with large femora from Koobi Fora. If it were paired with the smallest femora, its position would differ substantially. In contrast, there is a completely different body to brain size scaling relationships in *H. sapiens* than in *H. erectus* and the two samples are quite distinct. This difference in brain size–body size relationship in early *Homo* versus *H. sapiens* has previously been suggested by Walker (1993).

### 11.3.4 The influence of cranial size on expression of cranial characters

Given this vast range in cranial size (as proxied by brain size), it seems reasonable then to assess whether or not vault size affects any of the vault characters commonly considered in the definition of *H. erectus* such as supraorbital toral size, cranial thickness, cranial angulation, and so on. To assess this possibility, we look at the bivariate relationship between individual cranial capacity and cranial thickness (at parietal eminence, bregma, lambda, asterion, midfrontal), relative development of the occipital torus and supraorbital torus, angulation of the occipital, as well as with other cranial size variables (foramen magnum length, interorbital breadth). Modern human data are again from Brown (2005), Berschini and Haversat (1980), and also Oetteking (1930). The fossils included are all those from Africa or Asia that preserve the particular variable in question and the largest and smallest *H. habilis*. It is important to note that the following discussion relates only to relationships between our “independent” size variable (capacity) and that other significant scaling relationships may exist between other cranial variables. Our choice of capacity as an overall indicator of vault size is appropriate given our concern with calvarial characters.

Within the *H. erectus* sample, more than half the characters tested show significant correlations with cranial capacity (Table 11.8). Thus, some characters commonly argued as representative, particularly, of Asian *H. erectus*, most notably a thick supraorbital torus and thick posterior vault (at lambda, asterion,
and midoccipital), increase as a function of size (Figure 11.8). Other indicators of *H. erectus*, such as thickness at bregma (indicative of a bregmatic eminence), angulation of the occipital, and angulation of the petrous temporal, are not correlated with brain size and thus may be better taxonomic indicators. Foramen magnum size and interorbital breadth also scale with brain size.

When the largest and smallest *H. habilis* are added to the sample, existing correlations between vault size and cranial characters are strengthened and one additional parameter, parietal eminence thickness, is significantly correlated with size. As the plots suggest, on these indicators (posterior cranial thickness, supraorbital toral thickness, foramen magnum length, and interorbital breadth) it is difficult to distinguish smaller *H. erectus* from large *H. habilis*. And when

<table>
<thead>
<tr>
<th></th>
<th>H. erectus sample only</th>
<th>H. erectus and early Homo sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R²</td>
<td>Slope</td>
</tr>
<tr>
<td><strong>Face</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supraorbital torus height</td>
<td>0.250</td>
<td>0.649</td>
</tr>
<tr>
<td>Interorbital width</td>
<td>0.611</td>
<td>0.862</td>
</tr>
<tr>
<td><strong>Thickness</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At midfrontal</td>
<td>0.214</td>
<td>0.862</td>
</tr>
<tr>
<td>At bregma</td>
<td>0.023</td>
<td>0.138</td>
</tr>
<tr>
<td>At parietal eminence</td>
<td>0.242</td>
<td>0.879</td>
</tr>
<tr>
<td>At lambda</td>
<td>0.357</td>
<td>0.862</td>
</tr>
<tr>
<td>At external occipital protuberance</td>
<td>0.280</td>
<td>0.930</td>
</tr>
<tr>
<td>At asterion</td>
<td>0.509</td>
<td>1.608</td>
</tr>
<tr>
<td><strong>Size</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Occipital angulation</td>
<td>0.087</td>
<td>–0.208</td>
</tr>
<tr>
<td>Petrous angle</td>
<td>0.551</td>
<td>–0.702</td>
</tr>
<tr>
<td>Foramen magnum length</td>
<td>0.634</td>
<td>0.409</td>
</tr>
</tbody>
</table>

Table 11.8
Results of ordinary least-squares regression for cranial capacity and other cranial variables. ns = not significant.
evaluating specimens in this size range, other characters such as those on the cranial base should be used to guide taxonomic attributions. It should also be noted that in some measures (e.g., occipital thickness), smaller *H. habilis* is even thicker walled than comparably sized *H. erectus* specimens.

*H. sapiens* data are not available for all these dimensions, but where they are they show an interesting pattern of difference from *H. erectus* (Figure 11.9). Foramen magnum length scales with a similar slope as in *H. erectus*, with a grade shift of the regression line marking the absolutely larger cranial capacities of...
H. sapiens. However, interorbital breadth and cranial thickness measures show no discernible scaling relationship to cranial capacity within H. sapiens.

With regard to the recently discovered LB1 from Flores, sizes measures such as foramen magnum length plot as predicted for its capacity, however, cranial thickness data suggests that the inflammatory condition noted on the frontal of this individual (Brown et al. 2004) is likely to have affected some of its cranial thickness values as it plots substantially too thick for its cranial size as compared to the fossil Homo regression. Indeed, only the thinner regions of its vault (lambda, parietal eminence) plot at or near where predicted for its cranial size. In analyses of LB 1, the revised cranial capacity of 417 cm³ found by Falk et al. (2005) is used.

These analyses clearly suggest that characters relied upon in the past to identify H. erectus or to distinguish Asian from African morphs, notably overall large size, thick vaults, and large brows are merely the result of predominantly large brain sizes in the earliest discovered specimens. The smaller sizes of more recently discovered African and West Asian specimens result in the smaller scale of these prototypical characters. The current data do not then support an African–Asian split on these characters.
What then is *Homo erectus*?

In the end, we are left with the problem of defining the taxon. The pre-*H. sapiens* lineage appears, at least in some characters, to be scaled in terms of expression of certain cranial features previously used to define *H. erectus* (Figure 11.8). Principally, we can no longer assert that *H. erectus* is either absolutely or relatively thicker walled than other earlier taxa. The bregmatic eminence appears to be an exception to this rule. Similarly, the difference in supraorbital toral thickness between most Asian and African specimens is in part attributable to overall size. Likewise, absolutely and relatively larger brain sizes are not present in all members of the lineage. Although the data are more sparse, body size appears also to be scaled in presapiens *Homo* suggesting that while some populations of *H. erectus* may have been the first to attain these large sizes, it is a difference in degree not kind from earlier species (Figures 11.6 and 11.7). And large body size was not characteristic of all *H. erectus* populations at all times and all places, suggesting the need to be more sensitive to the potential of local adaptation in *H. erectus* populations. Indeed, there is some evidence of clinal variation in body size in *H. erectus* along climatic lines. However, higher latitude *H. erectus* do not appear to have approached the larger body size maxima seen in Middle and Late Pleistocene archaic *Homo* (Ruff et al. 1997).

At this point, it may be appropriate also to discuss the uneven nature of the samples with which we work. The depositional contexts of both Dmanisi and Ngandong that have yielded relatively large assemblages suggest relatively short depositional durations, perhaps also indicating a closer genetic relationship among individuals within these sites than say among the various Koobi Fora or Sangiran individuals. Indeed, the Dmanisi sample also yields several examples of dental agenesis and malrotation that may indicate some degree of inbreeding (LB 1 shows similar malrotation; Brown et al. 2004). We do not have a solution to dealing with this unevenness but point to it as a place of concern and interest for further exploration.

Based on these conclusions regarding the scaling of characters among *H. erectus*, the distinction between *H. erectus* and *H. ergaster* seems less justifiable than it has in the past. However, this geographically diverse, temporally transgressive sample does appear to separate from *H. sapiens* particularly in regard to the body size–brain size relationship and the transposition of some scaling relationships in the head relative to *H. erectus* due to large brain size (Figures 11.6 and 11.7). Likewise, based on these preliminary analyses, some scaling relationships, for example between interorbital breadth and capacity, seem to be wholly different from *H. erectus*. Additionally, scaling relationships that are present in *H. erectus*, such as those between cranial capacity and parietal
eminence thickness, EOP thickness and lambda thickness do not exist in H. sapiens (Figure 11.9).

The distinction between H. erectus and H. habilis is murkier than between H. sapiens and H. habilis. As mentioned, H. habilis and H. erectus tend to show similar scaling relationships for the features that we have considered. Whether H. habilis and H. erectus can be distinguished once size is taken into account remains to be explored further.

Acknowledgments

We are grateful to the editors for their invitation to join this volume and to our many colleagues worldwide for access to specimens and intriguing discussion. We are particularly grateful to E. Indriati, M. Leakey, L. Leakey, W.R. Leonard, and A. Taylor for general support and discussion.

References


Black D (1931) On an adolescent skull of Sinanthropus pekinensis in comparison with an adult skull of the same species and with other hominid skulls, recent and fossil. Palaeontol Sin Ser D 7(2): 1–114


Dawson C, Smith Woodward A (1914) Supplementary note on the discovery of a Paleolithic human skull and mandible in a flint bearing gravel overlying the Wealdon (Hasting Beds) at Piltdown, Fletching (Sussex). Q J Geol Soc London 70: 82–90


De Mendonça MC (2000) Estimation of height from the length of long bones in a Portuguese
Dubois E (1894) Pithecanthropus erectus eine menschenaehnlich Uebergangsform aus Java. Landsdruckerei, Batavia
Dubois E (1924) On the principal characters of the cranium and the brain, the mandible and the teeth of Pithecanthropus erectus. Proc Royal Acad Amsterdam 27: 265–278
Dubois E (1926b) Figures of the femur of Pithecanthropus erectus. Proc Royal Acad Amsterdam 29: 730–743
Dubois E (1932) The distinct organization of Pithecanthropus of which the femur bears evidence, now confirmed from other individuals of the described species. Proc Royal Acad Amsterdam 35: 716–722


Jacob T (1967) Some problems pertaining to the racial history of the Indonesian region. Utrecht, Drukkerij Neerlandia


Krogman WM (1978) The human skeleton in forensic medicine, 3rd printing. C.C. Thomas, Springfield IL


Defining Homo erectus: size considered
Oetteking B (1930) Craniole of the North Pacific Coast. E. J. Brill Ltd. New York
Pearson K (1898) On the reconstruction of the stature of prehistoric races. Philosophical transactions of the royal society 192A: 169–244
studies of an extinct human species. Cambridge University Press, New York
Tobias PV (1995) The place of Homo erectus in nature with a critique of the cladistic
Weidenreich F (1941) The extremity bones of Sinanthropus pekinensis. Palaeontol Sin NS D 5: 1–150
(ed) 100 Years of Pithecanthropus, the Homo erectus problem. Cour Forsch Inst Senckenberg 171: 159–165
12 Later Middle Pleistocene

Homo

G. Philip Rightmire

Abstract

Hominin fossils are known from Middle Pleistocene localities in Africa, Europe, South Asia, and the Far East. It is recognized that these individuals display traits that are derived in comparison to the condition in Homo erectus. However, the skulls retain numerous primitive features that set them apart from modern humans. Faces are massively built with strong supraorbital tori, frontals are flattened, and vaults remain low with less parietal expansion than in Homo sapiens. The hominins from Bodo, Broken Hill, and Elandsfontein in Africa are quite similar to their Middle Pleistocene contemporaries in Europe. Crania and jaws from Arago Cave and Petralona, and the spectacular assemblage from Sima de los Huesos, are particularly informative. In sum, this evidence suggests a speciation event in which H. erectus gave rise to a daughter lineage. At or before the beginning of the Middle Pleistocene, new populations spread through western Eurasia and perhaps also to the Far East. How the fossils should be treated taxonomically is currently uncertain. One view recognizes two species in Europe (in addition to the later Neanderthals), both of which are distinct from mid-Pleistocene Africans. Another perspective notes that differences between the specimens can be attributed to geography, time, or intragroup variation. Most if not all of the fossils can be accommodated in one species. If the Mauer mandible is included in this hypodigm, then the appropriate name is Homo heidelbergensis. This species is probably ancestral to both the Neanderthals in Europe and the earliest representatives of H. sapiens in Africa.

12.1 Introduction

Humans evolved in Africa and were confined to that continent for much of their early history. The first dispersals from Africa into Eurasia occurred shortly after 2 million years ago (Ma). These migrants were probably representatives of Homo erectus (sometimes called Homo ergaster). Traces left by these hominins have been recovered from the site of ‘Ubeidiya in the central Jordan Valley and at Dmanisi in...
the Georgian Caucasus. Such early occupations were likely transitory and did not result in permanent settlements. However, groups of *H. erectus* were able to travel relatively quickly across southern Asia to the Far East, where they were established both in Java and in China by about 1.6 Ma. The first penetration westward into Europe apparently came much later. There are indications that humans were moving into the Mediterranean region after 1 Ma, but the initial populating of Europe north of the major mountain barriers is documented only after 500 thousand years ago (ka). The biological identity of the first Europeans is unclear, but it is agreed that these hominins differ from *H. erectus*. Many of the ancient fossils are presently assigned to the species *H. heidelbergensis* (named originally from a mandible found near Heidelberg in Germany). *H. heidelbergensis* or perhaps other closely related species are known also from Middle Pleistocene localities in Asia and Africa ([Figure 12.1](#)). These people seem to have been more advanced in behavior than their predecessors, and there is evidence that

![Figure 12.1](#)

Map giving the locations of Middle Pleistocene localities where important hominin fossils have been discovered
12.2 The Middle Pleistocene of Africa

In Africa, fossils from the early Middle Pleistocene are clearly different from *H. erectus* in cranial capacity (approximately equal to brain size), width of the frontal bone, proportions of the occipital region, and anatomy of the underside of the skull. Where it is preserved, the face is still heavily constructed, but the brows, nasal profile, and bony palate more closely resemble the condition seen in later humans. In many instances, the hominins are found with stone tools that are more carefully shaped than the choppers and relatively crude hand axes associated with *H. erectus*. From Bodo in Ethiopia to Elandsfontein in South Africa, a shift toward the manufacture of thinner, more finely flaked bifacial tools is documented in the Middle Pleistocene, and it is reasonable to link this change in behavior to a speciation event in which *H. erectus* gave rise to a daughter lineage exhibiting increased relative brain size (encephalization).

12.2.1 Bodo

One important specimen came to light in 1976 at Bodo, in the Middle Awash region of Ethiopia (*Figure 12.1*). The Bodo cranium, and later a broken parietal from a second individual, were found in conglomerates and sands containing mammalian bones and Acheulean tools (Kalb et al. 1980; Clark and Schick 2000; Gilbert et al. 2000). Fauna from the Bodo site has been compared to that from Bed IV at Olduvai Gorge and Olorgesailie in Kenya, and an early Middle Pleistocene date is indicated. $^{40}$Ar/$^{39}$Ar measurements reported by Clark et al. (1994) support this biochronology, and the evidence points to an age of about 600 ka for the Bodo hominins.

The face and the anterior part of the Bodo braincase are preserved (*Figure 12.2*). There are some cut marks on the facial bones and these indicate intentional postmortem defleshing, as documented by White (1986). It can be established that Bodo is like *H. erectus* in some features. The massive facial bones, projecting brow, low frontal with midline keeling, parietal angular torus, and thick vault give the specimen a pronounced archaic appearance. In other respects, the cranium is more specialized (derived) in its morphology. Brain size is close to
1,250 cm$^3$ and is thus substantially greater than expected for *H. erectus*. Frontal bone proportions, the high-arched shape of the squamous temporal, and some traits of the cranial base are like those of more modern humans. Although the face is very broad and heavily constructed, the supraorbital tori are divided into medial and lateral segments, the margin of the nose is vertical rather than forward sloping, and the incisive canal opens into the front of the hard palate (Rightmire 1996). These are derived (apomorphic) conditions present in the face of recent *Homo*.

### 12.2.2 Broken Hill and Elandsfontein

Another African specimen is the cranium from Broken Hill (now Kabwe) in Zambia, discovered by miners in 1921. Quarrying for lead and zinc ore had already removed most of a small hill, when the miners broke into the lower part of an extensive cavern. Published reports do not all agree on this point, but apparently the cranium was picked up by itself, not in clear association with other hominin remains. The fossil is in remarkably good condition. The face is massive, with some of the heaviest brows on record. The frontal is flattened with slight midline keeling, and the vault is low in profile. Shortly after it was found, the fossil was attributed to the (new) species *H. rhodesiensis* (Woodward 1921). In its overall morphology, however, Broken Hill resembles *H. erectus*, and indeed, it has been classified this way on more than one occasion. At the same time, there are

![Figure 12.2](image12.png) Facial and oblique views of the cranium from Bodo, Ethiopia. The projecting glabellar region, wide interorbital pillar and massive zygomatic (cheek) bones give the face an archaic appearance similar to that of *H. erectus*. Other traits including the vertical border of the nasal aperture are interpreted as apomorphies shared with later humans.
apomorphic features shared with later humans. The temporal squama is high and arch-shaped, and the upper scale of the occipital is expanded relative to its lower nuchal portion (where the neck muscles are attached). Several discrete characters of the temporomandibular joint region are specialized. These include a raised articular tubercle and a sphenoid spine. More changes are apparent in the face, where the lateral border of the nasal aperture is set vertically, and the palatal anatomy is like that of later people (Rightmire 2001).

Another cranium quite similar to that from Broken Hill comes from the farm Elandsfontein, near Saldanha Bay on the Atlantic coast of South Africa. At Elandsfontein, there is an expanse of sandveld that has long been a focus of attention for paleontologists. Dunes migrate across this area, and in between the dunes, there are swales resulting from deflation. Whether the ancient horizons exposed in these “bays” are stratified land surfaces or simply mark the (seasonal) fluctuations of the water table is unclear. Given either of these interpretations, it is evident that during the mid-Quaternary, the region supported wetlands and water holes, with plenty of grass (Deacon 1998). Animals, many of them bovids or other large herbivores, were attracted to the water. The fauna includes numerous archaic elements such as a sabertoothed cat, a sivathere, and a giant buffalo. Altogether, some 18 of 45 mammalian species collected at the site are extinct. Comparisons conducted by Klein and Cruz-Uribe (1991) imply that the bones were accumulated between 700 and 400ka.

Much of the work at Elandsfontein has been surface prospecting, and it was during one such visit in 1953 that investigators picked up pieces of a human skullcap. The reconstructed Elandsfontein cranium is composed of the frontal and parietal walls and some of the occiput. The bones are cracked and heavily weathered, but the braincase is not distorted. There are some similarities to *H. erectus*, but certainly the better match is with Broken Hill. These two Middle Pleistocene specimens are alike not only in overall proportions but also in many anatomical details. The Elandsfontein brow is almost as thick as that of Broken Hill, and the frontal contours are the same. Radiographs show that the frontal sinus is large and complex, reaching well up into the squama in both cases (Seidler et al. 1997; Rightmire unpublished observations). The South African frontal bone gives a breadth index of 91.9 and is thus slightly less constricted than that of Broken Hill, for which the ratio of least width to greatest breadth is 83.0. Sagittal and coronal measurements of the parietal are similar in the two individuals as is the length and orientation of the upper scale of the occipital. Unfortunately, the Elandsfontein base is missing, and there is no face. These are just the regions where one would expect to find additional apomorphies setting the South African hominin apart from *H. erectus*. 
12.2.3 Lake Ndutu

A fourth Middle Pleistocene specimen is known from Lake N’dutu. This seasonal soda lake is located at the western end of the Main Gorge at Olduvai, in northern Tanzania. Excavations conducted near the lake margin in 1973 produced an encrusted human cranium, along with other fossils and numerous artifacts (Mturi 1976). Initially, the stone assemblage included mostly spheroids, cores, and flakes, but hand axes were picked up during later visits to the site. All of this material is thought to be derived from archeological horizons in a greenish sandy clay, tentatively correlated with the upper Masek Beds at Olduvai.

When it was found, the cranium was severely damaged and encased in a clay matrix. The process of cleaning and reconstructing the fossil has been described by Clarke (1990). These efforts were generally successful, but the face is quite incomplete, as is the frontal bone. There are gaps in the parietals as well. The braincase is relatively small, with a capacity of only about 1,100cm³. Just a fragment of the supraorbital region is preserved, and the torus is projecting, if not especially thickened. Bossing of the parietals is emphasized in Clarke's reconstruction. This has perhaps been overdone with plaster, but the walls of the vault appear to be more convex than would be the case for *H. erectus*. Also, the upper plane of the occiput is vertical, above the moundlike transverse torus. The morphology of this torus is in keeping with other characters suggesting that Ndutu could be female, in comparison to males such as Bodo or Broken Hill.

12.2.4 Florisbad

Several additional fossils are more fragmentary and therefore somewhat less informative. An example is the cranium from spring deposits at Florisbad in South Africa, consisting only of facial parts, the frontal bone, and pieces of the parietals. Early studies compared the hominin to recent populations, but it is important to emphasize that Florisbad is far from modern in its morphology. Glabella (in the midline above the nasal root) is projecting, as is the brow on either side. The facial bones as repositioned by Clarke (1985) suggest that the nasal cavity is large and the cheek is flattened, without obvious infraorbital hollowing. The face is less heavily constructed than that of Broken Hill but otherwise not dissimilar. A human upper molar tooth from Florisbad has been dated by ESR to 259ka (Grün et al. 1996).
Several sites in the Omo region of southern Ethiopia, explored initially in 1967, have recently been revisited. Human remains are known from both the PHS and KHS localities in Member I of the Kibish Formation, now considered to be 200–100 ka in age (Assefa et al. 2000). Omo 2 is an isolated surface find from PHS, lacking archeological associations. This partial cranium is low in contour and decidedly massive in its construction, with a blunt frontal keel and a strongly angled occiput. Other likely primitive features include the shape of the deep mandibular cavity lacking any distinct articular tubercle and the absence of a sphenoid spine. Nevertheless, the vault is large overall. The frontal bone is broad and relatively unconstricted, and the parietal walls show some outward curvature (limited to the regions below the temporal lines). The supraorbital torus is extensively damaged, and none of the face is preserved.

Omo 1 was excavated at the KHS site, from which there is now a large collection of Middle Stone Age artifacts. This individual is represented by only small portions of a skull, but much more of the postcranial skeleton is present. The cranium as reconstructed by several workers is globular in form, with expanded parietals and an occipital that is more rounded than that of Omo 2. To the limited extent that these can be checked, cranial superstructures (crests and tori) are not strongly expressed. The anterior part of the mandible shows clear signs of chin formation. Given these important markers of modern morphology, there is general agreement that Omo 1 should be regarded as early \textit{H. sapiens}. However, there is a question as to whether the Omo 1 skeleton can be grouped with the more archaic Omo 2 remains, or whether these individuals should be placed in different populations. The PHS and KHS localities are some distance apart, but geological evidence indicates that both lie in Member I of the Kibish deposits. There is no basis for regarding Omo 2 as substantially more ancient than Omo 1, and the skeletons must be close to 195 ka (McDougall et al. 2005).

Specimens from Herto in the Middle Awash region confirm the presence of \textit{H. sapiens} in northeastern Africa late in the Middle Pleistocene. Three fossilized crania recovered in 1997 show cut marks associated with postmortem defleshing and are associated with a stone tool assemblage that can be characterized as late Acheulean or Middle Stone Age. The bones and artifacts are dated radioisotopically to between 160 and 154 ka (Clark et al. 2003). One of the adult crania (BOU-VP-16/1) is intact, with a brain size estimated as 1,450 cm³ (White et al. 2003). This individual is ruggedly built, with a very prominent, bilaterally arched glabella, a long vault, and a distinctly flexed occipital. The parietal walls are convex rather than inward sloping, and the index of neurocranial globularity...
(Lieberman et al. 2002) calculated as ca. 0.54 for BOU-VP-16/1 is high enough to be within the range expected for anatomically modern humans. A second adult cranium is less complete, and there is a child estimated as 6–7 years in age. As a group, the Herto individuals are very robust but display morphologies that place them close to recent populations. White et al. (2003) have referred the fossils to a new subspecies of *H. sapiens*.

### 12.3 Middle Pleistocene hominins from Europe

Skulls very similar to those from Africa have been found in western Eurasia. Several of the principal localities lie close to the Mediterranean Sea, but it is apparent that humans were also able to reach Britain and central Europe, relatively early in the Middle Pleistocene (Figure 12.1).

#### 12.3.1 The cranium from Petralona

Petralona lies near the city of Thessaloniki in northern Greece. The exact provenience of the hominin fossil found within cave deposits containing the bones of numerous extinct animals is uncertain, but the Middle Pleistocene antiquity of this material is not in doubt. The cranium itself is exceptionally well preserved (Figure 12.3) and would have enclosed a brain close to 1,230 cm$^3$ in volume.
(Stringer et al. 1979). Supraorbital tori are about as massive and projecting as in Broken Hill, while CT scans show that the frontal sinuses are greatly expanded. These air cavities extend posteriorly toward bregma and also laterally, where they are separated from the sphenoid sinuses only by thin bony partitions (Seidler et al. 1997). The frontal bone itself is relatively shorter and broader than in Broken Hill. The ratio of least to greatest frontal breadths is 91.6; postorbital constriction is thus less pronounced than in Broken Hill but comparable to that estimated for the Elandsfontein specimen. Petralona also differs from Broken Hill in having a wider cranial base and a less prominent torus crossing the occipital bone. However, the two hominins are alike in many other aspects of vault shape, in height, breadth, and massive construction of the upper face and cheek, and in several measures of facial projection.

### 12.3.2 Arago Cave

Much the same conclusion applies to the less complete cranium from Arago Cave in France dated to about 450 ka. The partial cranium numbered Arago 21 has a face that is largely intact but damaged as a result of its long interment in compacted cave sediments. The frontal bone, interorbital pillar, nose, and cheeks show numerous cracks, and areas of localized crushing are present. The discoverers have been able to correct some of this damage in a reconstruction, but significant distortion remains. Nevertheless, it is evident that Arago 21 is somewhat smaller than Petralona or Broken Hill in brow thickness, upper facial width, and facial length. Height of the bony orbit and the subnasal part of the maxilla are especially reduced, and the nasal saddle seems to be less elevated relative to the orbital margins. Apart from these differences, Arago 21 is similar in its proportions to the Broken Hill cranium from Africa (Rightmire 2001).

Some workers discern resemblances to Neanderthals. Hublin (1996) and Arsuaga et al. (1997) note that the infraorbital surface of the Arago 21 maxilla is flattened and the cheek bones are obliquely oriented, as in Neanderthals. Also, there is forward protrusion of the face at subspinale (in the midline, just below the nasal opening), and the nasal aperture is bounded inferiorly by a sharp rim. These observations must be tempered by the fact that cracking and plastic deformation make it difficult to assess key aspects of morphology. The wall of the Arago 21 maxilla is generally flattened or even inflated in the manner characteristic of Neanderthals, but the cheek is slightly hollowed laterally, below the orbit. This feature cannot be due entirely to damage. Also, it is not clear
that the zygomatic bone is swept back (obliquely oriented) so noticeably as in later European populations. In facial forwardness at subspinale [as measured by the zygomaxillary angle of Howells (1973)], Arago 21 at 113° is in the Neanderthal range, and Petralona at 118° shows almost as much protrusion. But the value for Broken Hill is only 116°, so a low zygomaxillary angle does not align Arago 21 and Petralona with Neanderthals rather than with other Middle Pleistocene specimens. The sharp inferior margin of the Arago nose is indeed reminiscent of that in Neanderthals. However, there is variation in this feature. Petralona is rather less like the Neanderthals, while some later Europeans including the Sima hominins (Section 12.3.3) have a pattern of cresting on the nasal floor resembling that in Broken Hill or Bodo.

In addition to the partial cranium, the cave at Arago has yielded two mandibles. Arago 2 is the more complete, missing only the angle and ascending portion from the left side. This specimen has sustained damage anteriorly, where the symphysis and left corpus are cracked. Arago 13 is a large hemimandible (right side), in relatively good condition. Both specimens present a mix of archaic and more modern characters. Development of the lateral prominences, marginal tori and tubercles, and internal symphyseal buttresses is comparable to that observed in *H. erectus*, although the alveolar planum is steeper and less shelflike in the Arago individuals. Arago 2 displays definite incurving of the symphyseal face below the alveolar border. Here, the elements of a mental trigone are present, while in Arago 13, signs of “chin” formation are less clear. Both jaws have retromolar fossae. However, in Arago 13 this fossa is restricted, and the crown of M3 is partly obscured by the leading edge of the ramus when the specimen is viewed from the side.

The Arago mandibles are important not only because they reveal information about a Middle Pleistocene hominin population but also because they can be compared to the jaw from Mauer, near Heidelberg in Germany. Assigned a tentative age of ca. 500 ka, the Mauer fossil is likely to be one of the oldest recovered in Europe. It has often been described as primitive, with a massive body and very thick symphysis lacking any mental eminence. At the same time, the broad ramus, increased symphyseal height and moderate size of the teeth suggest a morphological pattern different from that of *H. erectus*. The mandible was referred to the (new) species *H. heidelbergensis* by Schoetensack (1908). As the Arago jaws resemble the Mauer specimen, it is possible to link the French assemblage with the same taxon. Similarities of the Arago 21 face to Petralona (or Broken Hill) in turn provide a formal basis for including other European (or African) individuals in *H. heidelbergensis*.
12.3.3 Sima de los Huesos, Atapuerca

The species *H. heidelbergensis* is increasingly well documented by the spectacular finds from Atapuerca in northern Spain. Excavations in the Sima de los Huesos have produced hominin remains representing virtually all parts of the skeleton. In addition to skulls, there are many postcranial bones, and it is clear that at least some of the Sima (male) individuals were tall and robust (Arsuaga et al. 1999). Somewhat surprisingly, sexual dimorphism is comparable to that expressed in recent populations. The cave also contains the bones of bears and a few other carnivores, but there are no herbivores that might represent food waste. With one exception, there are no stone artifacts. A single hand axe fashioned from red quartzite was discovered in 1998. Investigators working at the Sima have argued that the skeletons were deposited in this pit by other humans and that the unique hand axe documents symbolic behavior (The First Europeans 2003). Application of U-series dating to a speleothem present in the lower part of the stratigraphic sequence in the site suggests a date of >350 ka (Bischoff et al. 2003). An age of ca. 400 ka for the assemblage seems reasonable.

Two of the Sima adults provide estimates for brain size. At close to 1,100 cm$^3$, SH 5 is rather small, but SH 4 with a capacity of 1,390 cm$^3$ is one of the largest of all Middle Pleistocene specimens. The crania are primitive in some respects, and the massive face of SH 5 is surmounted by a prominent browridge. Vault bones are thickened, and both sagittal keeling and an angular torus are variably developed. The braincase is broadest in the supramastoid region or just above the ear openings. As do their European and African contemporaries, the Sima hominins also exhibit derived traits in the face, shape of the squamous temporal, proportions of the occipital bone, and structure of the cranial base.

An important question is the extent to which these people resemble the later Neanderthals of Europe. As described by Arsuaga et al. (1997), the midface of SH 5 seems to anticipate the distinctive morphology associated with Late Pleistocene Europeans. The infraorbital surface and the side wall of the nose meet at a shallow angle, so as to produce a slight concavity. The cheek region is thus not “inflated” in the extreme manner of Neanderthals, but it can be interpreted as intermediate in form. Also in the Sima sample, continuity of the supraorbital tori at glabella is said to be reminiscent of Neanderthals, and the broad nasal bones are set in a relatively horizontal orientation. At the rear of the cranium, the suprainiac area is large but not very depressed. This trait and the shape of the occipital torus may also foreshadow the Neanderthal condition. How these features are evaluated (whether any of them can be judged to be true Neanderthal apomorphies) will
determine how the Sima hominins as well as Arago and Petralona are related to populations outside of Europe and how these regional paleodemes should be treated in phylogenetic schemes.

### 12.3.4 The TD6 assemblage from Gran Dolina, Atapuerca

Additional evidence bearing directly on the first peopling of Europe is accumulating from another site in the Atapuerca region. Excavations at Gran Dolina have uncovered stone core-choppers and flakes, animal bones, and human remains dating to the end of the Early Pleistocene. An age slightly in excess of 780 ka for the TD6 level containing the fossils now seems to be established (Falgue`res et al. 1999). Cranial specimens include a juvenile face, an adult cheek bone, part of a subadult frontal with some of the brow, and a piece of the cranial base on which most of the joint cavity for the mandible is preserved. There are also broken lower jaws with teeth, along with vertebrae, ribs, and bones of the hand and foot.

Arsuaga et al. (1999) argue that the TD6 people are not *H. erectus*. Morphology of the hollowed cheek region, vertical orientation of the nasal aperture, features of the hard palate, form of the developing (but already substantially thickened) brow, a wide frontal, the shape of the temporal bone at the side of the vault, and the apparently modern mandibular joint, all suggest that the Gran Dolina fossils are different from *H. erectus* and more like later humans. Also, there can be little doubt that this population is distinct from the later Neanderthals. The hollowed cheek (bearing a “canine fossa”) points toward this conclusion, and neither in the juvenile nor in the adult faces is there much sign of the specialized Neanderthal condition. One partial mandible is generalized in its morphology, while the teeth resemble those of European and African Middle Pleistocene hominins.

Given this complex of traits, the Gran Dolina material may represent a new species. The name *H. antecessor* was proposed by Bermúdez de Castro et al. (1997). However, the number of fossils is still quite small, and several of the craniodental remains are fragmentary and/or subadult. A fair question is whether there is presently enough evidence to separate the TD6 assemblage from other penecontemporary fossils already on record. In particular, it must be asked whether the Gran Dolina bones and teeth differ from those of other early Europeans such as Mauer, Arago, and the Sima de los Huesos. Much attention has been focused on the development of a “canine fossa” in the midface. Hollowing is indeed apparent in the cheek of the TD6 juvenile, but a fossa is less obvious in the TD6 adult. This feature is variable in its expression in other populations, and the significance of this pattern is unclear. In the mandible, teeth, and postcranial
bones, there seem to be few traits that differentiate the Gran Dolina hominins from Europeans of the Middle Pleistocene.

**12.4 South Asia and the Far East**

One South Asian locality deserving mention is the Narmada Valley in central India (Figure 12.1). Part of a cranium was found there in 1982, embedded in a conglomerate containing animal bones and a scattering of Acheulean artifacts. Dates for this material are poorly constrained, but it is probably of Middle Pleistocene age (Sonakia and Biswas 1998). Unfortunately the skull is damaged and lacks most of the face. Narmada has been described by its finders as *H. erectus*, but it is better compared to *H. heidelbergensis* (Kennedy et al. 1991). In its overall morphology, the cranial vault is not very different from the African and European hominins already discussed.

Early humans occupied China before 1.6 Ma (Zhu et al. 2004). This part of Asia has been a focus of research in paleoanthropology for quite a long time. Apart from the famous discoveries of *H. erectus* at Zhoukoudian, there are important sites dating to the later Middle Pleistocene. One is Dali and another Jinniushan, both of them in northern China. The Dali cranium was found in river terrace deposits with stone flakes and fauna. The Jinniushan skeleton was recovered from cave fill containing animal bones but no artifacts. ESR and U-series dates obtained from animal teeth suggest ages of perhaps 300–200 ka.

Dali is much of a cranium, damaged on the right side and at the base. The alveolar process and palate have been crushed upward. The specimen is otherwise undistorted and carries a lot of information. It has most often been described as “archaic” *H. sapiens*, intermediate in form between *H. erectus* and recent humans. Indeed, there are similarities to *erectus*, and these include the heavy brow, a long low vault that is broad across the base, and the sharply angled occiput. The temporomandibular joint cavity is offset laterally, and the cranial bones are thickened. These traits are best described as primitive retentions. At the same time, Dali exhibits other advanced features that link it to later populations. There is not much postorbital constriction, and the parietal walls are vertical rather than inward sloping. Both the high temporal squama and the proportions of the occiput depart from the *erectus* condition. In the facial skeleton, there are other changes in nasal morphology and to the palate.

The Jinniushan cranium has been reconstructed several times, and there are gaps in the face, the frontal region, and the base. The brow is somewhat less massive than in Dali, but there is an eminence behind bregma, and the occiput is flexed. In other respects, the specimen differs from *H. erectus*. Brain volume is
close to 1,300 cm³. The border of the nasal aperture is vertical (rather than angled forward), and the nasal sill is crested. On the palate, the incisive canal opens anteriorly (just behind the incisor roots) as in recent humans.

In many anatomical details, both Dali and Jinniushan are like other Middle Pleistocene hominins from Africa or Europe. Comparisons based on facial measurements show that the Chinese specimens resemble Broken Hill to about the same extent as does Arago 21 (Rightmire 2001). There are some differences relative to Broken Hill, particularly in upper facial height (reduced in Dali and Jinniushan) and flattening below the nose (more pronounced in Jinniushan). Also, the Dali cheek exhibits a “canine fossa.” This feature has been taken as a basis for regarding the Chinese fossil(s) as distinct from western populations, but in fact hollowing of the infraorbital surface can be documented for faces outside of the Far East. Finds from Gran Dolina suggest that this feature may appear in Europe at the beginning of the Middle Pleistocene (Section 12.3.4). The recognition of such variation will make it harder to argue for isolation of the major Old World geographic provinces.

12.5 Brain size and encephalization

Many of the Middle Pleistocene hominins have brains that are enlarged relative to those of *H. erectus*. For 10 of the more complete crania including Bodo, Broken Hill, Petralona, two of the Sima de los Huesos adults, Dali and Jinniushan, average capacity is 1,206 cm³. For 30 *H. erectus* individuals, the mean volume is only 973 cm³. This difference is substantial, and it can be determined that a number of the Middle Pleistocene specimens actually lie beyond the limits predicted for an average *H. erectus* of comparable antiquity. That is, the observed increase is not simply an “expected” result of gradual evolution occurring within a single lineage (Rightmire 2004). Further support for this contention is provided by estimates of relative brain size. An index constructed by dividing the cube root of endocranial volume by orbital height averages 3.04 for eight of the more intact Middle Pleistocene skulls. This figure is significantly greater than the mean index of 2.68 calculated for *H. erectus*, and it demonstrates that change in cranial capacity is not attributable solely to an increase in body mass of the mid-Quaternary hominins.

Encephalization quotients (EQ) can also be obtained for a number of the specimens. This entails first estimating body mass from orbital height (following Aiello and Wood 1994) and then deriving EQ from the relationship of brain weight to body mass established for mammals by Martin (1981). Here, there are various complications. Apart from the error associated with any weight
estimate, there is the fact that the regression equations of Aiello and Wood (1994) are based on several species. Because EQ is a function of body mass predicted for individuals using an interspecific equation, comparisons of the EQ values determined for fossils may be misleading (Smith 2002). It is thus possible that EQ will be less useful than the relative brain size index, which is based directly on orbit height (treated as a proxy for body size). In any case, six *H. erectus* crania from Africa and Asia are complete enough to supply the necessary measurements, and the average EQ is 3.61 (Rightmire 2004). This result is comparable to that reported by Ruff et al. (1997), who employ mean estimates of brain and postcranially based body masses to compute EQ values of 3.40 and 3.46 for temporally defined (Early Pleistocene to early Middle Pleistocene) assemblages.

During the balance of the Middle Pleistocene, a rise in EQ is apparent. Bodo and Broken Hill remain within the range observed for *erectus*, but other individuals have higher values and the average for eight specimens is 5.26. The magnitude of this increase is greater than that determined by Ruff et al. (1997) for humans of mid-Quaternary age. These authors use unmatched brain and body weights (means for samples of disassociated crania and postcrania) as a basis for their EQ calculations, and this may account for some of the difference in results. Also, orbit height may tend to underestimate body mass in comparison to predictor variables drawn from the postcranial skeleton. Nevertheless, there is evidence for a shift in brain size at or just before the onset of the Middle Pleistocene. This increase in both gross volume and encephalization seems to be linked to an episode of species change.

12.6 Phylogenetic hypotheses

Discoveries of new fossils, reassessments of specimens found earlier, and advances in the application of dating techniques show that hominins differing from *H. erectus* appeared in southern Europe before 780 ka and in Africa at about the same time. One reading of the record suggests that these European and African groups share a number of derived features (frontal proportions, parietal convexity, arching of the temporal squama, greater relative length of the occipital upper scale) probably tied to expansion of the brain and cranial vault. Other similarities to later humans are apparent in the facial skeleton (orientation of the nasal aperture, location of the palatal incisive canal) and perhaps the mandible (symphyseal height increased relative to the posterior corpus, incipient mental eminence). Postcranial bones known principally from the Sima de los Huesos in Spain suggest that the European hominins were heavily built, perhaps reflecting adaptation of body form to a temperate environment. In sum, the anatomical
evidence can be interpreted as supporting a claim that all of the earlier Middle Pleistocene fossils belong to a single lineage (Figure 12.4a). This species can be called H. heidelbergensis. Later in the Middle Pleistocene, some populations dispersed northward within Europe, where they were subject to long episodes of extreme cold. During glacial advances and retreats occurring over several hundred thousand years ago, these hominins continued to adapt to harsher (cold/dry) conditions and evolved the specialized craniofacial characters and body build of the Neanderthals. In this same interval of time, other representatives of H. heidelbergensis in Africa were becoming more like modern humans. Fossil finds from Irhoud in Morocco, the Omo in southern Ethiopia, Herto in the Middle Awash region, and Laetoli in Tanzania document this evolutionary progression toward H. sapiens.
Alternatively, it can be argued that H. antecessor is the ancestor to all later humans (Figure 12.4b). This species is considered to be descended from (African) H. erectus (Bermúdez de Castro et al. 1997). Rather, soon after its first appearance in Spain, H. antecessor must have given rise to H. heidelbergensis. In this scenario, the heidelbergensis lineage was confined exclusively to Europe, where its members gradually acquired the large nose, more projecting facial skeleton, and other morphology of the Neanderthals. This is the accretion hypothesis of Dean et al. (1998). Also, H. antecessor is presumed to have evolved an African offshoot, represented at localities such as Bodo, Broken Hill (Kabwe), and Elandsfontein. Although these Middle Pleistocene hominins are acknowledged as morphologically similar to (perhaps even capable of exchanging genes with) their European contemporaries, they are not assigned to H. heidelbergensis. Instead, the African fossils are lumped in a separate species, for which the nomen H. rhodesiensis is available. Whether this taxonomic view can be accepted will depend largely on the outcome of excavations that are continuing in the TD6 levels at Gran Dolina. It will be important to expand the sample of fossils documenting the earliest European settlers.

Another question is whether the far eastern specimens can be accommodated within one of these systematic frameworks. The answer is a tentative yes, although the evidence is sparse. Dali and Jinniushan do share a number of apomorphic traits with the western hominins. But there are some differences, and the face has been a focus of contention. Dali has a short face, and this would be true even if damage to the maxilla were corrected. Jinniushan also has a short clivus (the subnasal portion of the maxilla), and it is oriented vertically. In Dali, there is hollowing of the cheek below the orbit, and such excavation is not present in the African crania. Much has been made of this facial morphology, but in fact there is individual variation (see Section 12.5). The significance of the Dali “canine fossa” should not be overemphasized. It is possible to argue that the later Middle Pleistocene hominins of China document an eastward excursion of H. heidelbergensis, where this species is taken to be the link between H. erectus and all later humans. Dating is not very firm, but probably fossils such as Dali and Jinniushan are younger than those in Africa. This may suggest that H. heidelbergensis was a late arrival in the eastern part of Asia.

12.7 Archeology and behavior in the Middle Pleistocene

Whether or not (all of) the earlier Middle Pleistocene paleodemes can be assigned to one lineage, it is becoming clear that the hominins were more encephalized
than *H. erectus*. Also, there is evidence from archeology that these people were developing new behavior. Later Acheulean artifacts are known from numerous African sites, including Bodo, Olorgesailie, Isimila, Lake Ndutu, the Cave of Hearths, Elandsfontein, and Duinefontein 2. In general, later Acheulean hand axes can be characterized as thinner, more symmetrical, and bearing many more flake scars than their earlier counterparts. In some sites, relatively small hand axes are accompanied by flake tools resembling those of the Middle Stone Age (Klein 2000). While it is dangerous to expect universal associations of *Homo* species with particular industrial traditions, informative patterns may be uncovered (Foley and Lahr 1997). In virtually all mid-Quaternary African contexts, where diagnostic human bones are found with later Acheulean artifacts, the maker is *H. heidelbergensis* (or *H. rhodesiensis*). One may conclude that this species was capable of producing a tool kit more sophisticated than that utilized routinely by *erectus*.

In western Eurasia, hominins equipped with Acheulean tools were present by the onset of the Middle Pleistocene (780 ka) at Gesher Benot Ya'aqov in Israel (Goren-Inbar et al. 2000). Farther to the west in Europe, there are no Acheulean sites from the beginning of the Middle Pleistocene, but Boxgrove in Britain is likely to be 500 ka in age. This locality has yielded thin, extensively flaked flint bifaces, along with bones of horses and rhinoceroses bearing cut marks. The animals may well have been hunted and butchered. In addition, there is the shaft of a human tibia. The dimensions of this bone at midshaft are large, and the Boxgrove individual was probably quite massive. This hominin has been attributed to *H. heidelbergensis* by Roberts et al. (1994). Signs of later Acheulean toolmakers are known from Torralba and Ambroña in Spain, where the artifacts are again found with large herbivores, including elephants and horses (Freeman 1994). Acheulean artifacts occur also at several sites in France and Italy. At Castel di Guido in central Italy, finely flaked bifacial tools were produced from elephant bone (Villa 1991). At some other earlier Middle Pleistocene localities including Arago Cave, the stone industries contain small chopping tools and flakes but no hand axes (De Lumley et al. 1984). The reasons for this difference are unclear, but the availability of suitable raw materials, the constraints imposed by different types of stone, and the context in which tools were manufactured must be considered, along with the possibility that distinct cultural behaviors or styles are represented.

An isolated but particularly significant example of the skills acquired by mid-Quaternary Europeans comes from Schöningen in Germany. Three carefully crafted wooden throwing spears have been uncovered near a former lake, where they are associated with flint tools and chips (Thieme 1997). Scattered through
the same horizon are the remains of numerous horses. Many of the bones are cut-marked, and some of the animals must have been processed for meat and marrow extraction (Roebroeks 2001). More convincingly than other early European assemblages, the Schöningen discovery points to systematic hunting of large animals. Stalking and killing of agile or dangerous prey requires experience and practice, and it is reasonable to hypothesize that the people were cooperating with one another in these efforts. Increased levels of social cooperation and exchange of knowledge would have become the norm. And if the hunters at Schöningen (also at sites such as Boxgrove and Arago) were able to obtain large amounts of meat, they would likely have shared or exchanged food with other groups, perhaps at established meeting places (Roebroeks 2001). Certainly our understanding of the behavior of the early Europeans remains quite incomplete, but it is apparent that bands of H. heidelbergensis were not only skilled at flaking stone but also capable of interacting regularly in the pursuit of game and other social activities.

Acknowledgments

For access to fossil materials in their care, I thank my friends and colleagues at The Natural History Museum in London, the Muséum National d’Histoire Naturelle in Paris, the University of Thessaloniki, the National Museum of Ethiopia, the National Museum of Tanzania, the National Museum in Bloemfontein, the Iziko Museums of Cape Town, the Institute for Vertebrate Paleontology and Paleoanthropology in Beijing and Peking University. Anne Hull helped with the illustrations. The Boise Fund, the Leakey Foundation, and the Eckler Fund of Binghamton University (SUNY) provided grant support.

References

Bermúdez de Castro JM, Arsuaga JL, Carbonell E, Rosas A, Martinez I, Mosquera M (1997) A hominid from the Lower Pleistocene of
Atapuerca, Spain: Possible ancestor to Neandertals and modern humans. Science 276: 1392–1395
Klein RG, Cruz-Uribe K (1991) The bovids from Elandsfontein, South Africa, and their
implications for the age, palaeoenvironment and origins of the site. Afr Archaeol Rev 9: 21–79


Schoetensack O (1908) Der Unterkiefer des Homo heidelbergensis aus den Sanden von Mauer bei Heidelberg. Ein Beitrag zur Paläontologie des Menschen, Englemann, Leipzig


The first Europeans: Treasures from the hills of Atapuerca (2003) Junta de Castilla y Leon, New York


13 Neanderthals and Their Contemporaries

Katerina Harvati

Abstract

Neanderthals are the group of fossil humans that inhabited Western Eurasia from the mid Middle Pleistocene until approximately 30 thousand years ago (ka), when they disappear from the fossil record, only a few millennia after the first modern humans appear in Europe. They are characterized by a suite of morphological features, which in combination produce a unique morphotype. They are commonly associated with the Mousterian lithic industry, although toward the end of their tenure they are sometimes found with assemblages resembling those produced by early modern humans. Although there is still discussion over their taxonomic status and relationship with modern humans, it is now commonly recognized that they represent a distinct, Eurasian evolutionary lineage sharing a common ancestor with modern humans sometime in the early Middle Pleistocene. This lineage is thought to have been isolated from the rest of the Old World, probably due to the climatic conditions of the glacial cycles. Glacial climate conditions are often thought to have been at least in part responsible for the evolution of some of the distinctive Neanderthal morphology, although genetic drift was probably also very important. The causes for the Neanderthal extinction are not well understood. Worsening climate and competition with modern humans are implicated.

13.1 The discovery of Neanderthals: historical background

Although the first Neanderthal remains were discovered in the early nineteenth century (Engis child in 1830, Forbes Quarry adult in 1848), it was not until the discovery of the skeleton from the Neander valley in 1856, roughly coinciding with the publication of Darwin’s “The Origin of Species” in 1859, that the existence of an extinct kind of archaic humans was recognized. It is this locality that lends its name to the group, and it is there that the debate surrounding the relationship of Neanderthals with modern humans began.
The antiquity of the Neanderthal skeleton and its status as an archaic human was not immediately accepted. Instead, its peculiar anatomical attributes were considered the result of various pathologies, including rickets. Its antiquity was only firmly established with the eventual discovery of additional skeletons of similar morphology associated with lithic artifacts and extinct fauna.

Neanderthals were assigned to the species *H. neanderthalensis* as early as 1864 (King 1864). However, once their status as archaic predecessors of modern humans was accepted, their relationship with modern humans, and particularly modern Europeans, began to be intensely debated. The predominant view in the 1910s and 1920s was represented by scientists like Marcellin Boule and Sir Arthur Keith, who were among the most influential scholars of their day. They placed Neanderthals in their own species and rejected any ancestral role for them in the evolution of modern people, pointing out their “primitiveness” and presumed inferiority (Boule 1911–1913; Boule and Vallois 1957).

This perception of Neanderthals began to change during the 1930s. A rearrangement and “pruning” of the tangled hominin taxonomy was undertaken in the 1940s and 1950s by Mayr, Simpson, and Dobzhansky, who placed Neanderthals and other Middle Pleistocene fossil specimens within our own species, *Homo sapiens*. According to this view, Neanderthals were thought to have evolved into modern people through anagenetic evolution (Trinkaus and Shipman 1992, 1993; Tattersall 2000). This view has been reexamined in more recent years, with new evidence coming from modern human and fossil genetic studies, the development of better dating techniques and new approaches to the analysis of morphology. Currently, Neanderthals are commonly viewed as a distinct, Western Eurasian evolutionary lineage, which probably did not contribute significantly to the evolution of modern people.

### 13.2 Chronological distribution

The first human occupation of Europe is now placed between 700 and 800 thousand years ago (ka). Early human skeletal remains have been recovered from the Southern European sites Gran Dolina (Atapuerca, Spain; Bermudez de Castro et al. 1997), and Ceprano (Italy; Manzi et al. 2001), dating to ca. 780–800 ka. Lithic artifacts dated to ~700 ka from Pakefield, England (Parfit et al. 2005) recently documented an early human presence also in Northern Europe. These early European populations are considered by some to have been ancestral to the later, Middle Pleistocene Europeans and to Neanderthals (Bermudez de Castro et al. 1997), but may also represent unsuccessful early episodes of colonization that ended in local extinctions.
The first appearance of Neanderthals in the fossil record is not clear-cut. Neanderthal-like features appear for the first time in a mosaic fashion in Middle Pleistocene European humans, as e.g., in the large assemblage from Sima de los Huesos (Atapuerca), Spain (Arsuaga et al. 1997), recently dated to between 400 and 600 ka (Bischoff et al. 2003). Different Neanderthal-like traits appear at different times and places and in different combinations, but their presence in the European Middle Pleistocene fossil specimens suggests that the latter were early representatives of the Neanderthal lineage (Figure 13.1).

Figure 13.1
Frontal (a) and lateral (b) views of the Middle Pleistocene cranium from Petralona, Greece, showing incipient Neanderthal morphology in its facial region (Courtesy of Eric Delson and © Eric Delson, Photo by K. Harvati)

A progressively stronger expression of Neanderthal morphology is perceived through time, with specimens from the late Middle Pleistocene/early Late Pleistocene (∼200–100 ka) showing clear, albeit still not fully expressed, Neanderthal morphology. The full suite of Neanderthal features appears with the “classic” Neanderthals, in the Late Pleistocene, dated from approximately 70 to 30 ka. This group includes among others the famous “Old Man” of La Chapelle-aux-Saints, as well as the type specimen from the Neander valley, Feldhofer 1.

The last date of appearance of Neanderthals is somewhat easier to pinpoint, with the most recent specimens dated to somewhat prior to 30 ka (Figure 13.2). Thus the disappearance of Neanderthals from the fossil record occurred a few millennia after the first appearance of modern humans in Europe around 40 ka. The latest known Neanderthal skeletal remains come from geographically
Figure 13.2
Map of the geographic distribution of Neanderthals. Several sites preserving Neanderthal and pre-Neanderthal skeletal remains are shown.
disparate sites. They are found in layer G1 at Vindija, Croatia, recently redated to ca. 32–33 ka through direct accelerator mass spectrometry (AMS) radiocarbon dating and with collagen ultrafiltration (Higham et al. 2006). Most Neanderthal fossils at this site come from the older G3 layer (Ahern et al. 2002, 2004). A Neanderthal mandible from Zafarraya, Spain, was radiocarbon dated to between 30 and 33 ka (Hublin et al. 1995). Late Neanderthal remains have been recovered from two sites in France, both associated with the Châtelperronian lithic industry (see below). The partial skeleton from Saint-Césaire was dated to ca. 36 ka using the thermoluminescence (TL) dating method (Mercier et al. 1991). A subadult temporal bone (Hublin et al. 1996), as well as several Neanderthal isolated teeth (Bailey and Hublin in press), from Layer Xb at Arcy-sur-Cure (Grotte-du-Renne), were dated by AMS radiocarbon to ca. 34 ka (David et al. 2001). The Neanderthal occupation of the portuguese site Figueina Brava has also been dated to as recently as 30–29 ka with radiocarbon and ESR dating (Antunes et al. 2000). Finally, a very late direct radiocarbon date for one of the Mezmaiskaya (Russia) Neanderthal infants (~29 ka; Ovchinnikov et al. 2000) is now thought to have resulted from modern carbon contamination. Recent work on this site has produced much older new ESR dates (~60–70 ka for the first infant, and ~40 ka for the second; Skinner et al. 2005).

13.3 Geographic distribution

Neanderthals are commonly thought of as European hominins, and Europe is often considered as their geographical area of origin, with specimens outside the continent representing later range expansions (Hublin 1998, 2000). Within Europe, Neanderthals range from Iberia to Russia and from the Mediterranean to Northern Europe. Outside of the continent, their presence has been documented in the Near East and in Western Asia as far east as Uzbekistan (Figure 13.2). Eastern Neanderthals are often juxtaposed with those from western Europe in that their morphology is mosaic in pattern and not fully “Neanderthal” (Vandermeersch 1989; Smith 1991; Rak 1993). In some of their features, Eastern Neanderthals show conditions that have sometimes been perceived as more modern (though not all specimens show the same conditions for the same features), leading some to question the Neanderthal identity of these fossils (Arensburg and Belfer-Cohen 1998). A more widely accepted view is that the weaker expression of Neanderthal traits in these specimens reflects primitive retentions rather than affinities with modern humans (Stringer 1990). Similarly, weakly expressed Neanderthal morphology is found in early Neanderthals, such
as those from Saccopastore, again interpreted as retentions of primitive morphology (Condemi 1992).

### 13.4 Morphology

Neanderthals are characterized by a suite of distinctive cranial, mandibular, dental, and postcranial anatomical features (Figures 13.3 and 13.4), some of which represent retentions of ancestral conditions but many of which are derived for this group. Primitive traits, shared with the common ancestor of both Neanderthals and modern humans, include their low and elongated crania, heavy brow
ridges, large faces with large nasal apertures, and the lack of a chin. Neanderthals share some derived features with modern humans, including enlarged brains, reduced prognathism, a weak occipital torus, and a longer and more rounded occipital. A list of proposed deriver traits is provided in Table 13.1.
A detailed morphological description of the various anatomical areas follows.

### 13.4.1 Cranium

#### 13.4.1.1 Face

The Neanderthal face is characterized by a heavy, double-arched supraorbital torus which does not show distinct elements and grades smoothly onto the frontal squama. The orbits are large and rounded. The nasal aperture is very large and broad. The nasal cavity is voluminous and displays large medial projections and a bilevel internal nasal floor. Neanderthals show pronounced...
midfacial prognathism, and an oblique and inflated infraorbital plate with no canine fossa, which obliquely recedes into the zygomatic bone. The inferior root of the zygomatic is oblique, and not sharply angled, and the zygomatic processes are elongated and thin. Internally, the maxillary sinuses are large, while the frontal sinus is expanded laterally to fill most of the supraorbital torus (up to midorbit) but does not extend upward into the frontal squama (Heim 1974, 1976; Stringer et al. 1984; Rak 1986; Trinkaus 1987, 2003; Schwartz and Tattersall 1996a; Arsuaga et al. 1997; Franciscus 1999, 2003).

Several alternative hypotheses have been proposed to account for the distinctive Neanderthal facial morphology. The large nasal aperture and associated structures have been proposed to relate to cold-climate adaptation and to function in warming and humidifying inspired air, as well as to dissipate heat (Coon 1962; Dean 1988). Another interpretation sees the Neanderthal facial features as biomechanical consequences of intense paramasticatory behavior evidenced by the unusual anterior tooth wear pattern exhibited in this fossil group (Heim 1976; Rak 1986; Trinkaus 1987). However, large noses and nasal cavities are characteristic of warm-climate populations among recent humans (Franciscus 2005) and several studies have rejected the proposed biomechanical advantages of the Neanderthal face (Antón 1994, 1996). A third interpretation considers Neanderthal facial morphology as primarily the result of stochastic processes (Hublin 1998; Franciscus 2003, 2005).

13.4.1.2 Vault

Neanderthals show a particularly flat and elongated vault in lateral profile, and an “en bombe,” rounded profile in posterior view, with the widest point at the mid-parietals. The occipital region shows a highly convex occipital scale, with a flattening above lambda, termed occipital “bun” or “chignon.” The occipital torus is weak with no external occipital protruberance. It is inferiorly undercut by the nuchal plane, but not clearly defined superiorly, and shows a pitted oval depression above it (suprainiac fossa). The temporal squama is superoinferiorly low, anteroposteriorly short and symmetrically arched. The external auditory meatus is elevated relative to the zygomatic process and the floor of the glenoid fossa. The parietomastoid suture is relatively long and straight (Boule 1911–1913; Hublin 1978, 1988a, b, 1998; Stringer et al. 1984; Condemi 1988; Lieberman 1995; Schwartz and Tattersall 1996b; Dean et al. 1998; Harvati 2003b).

Neanderthal endocasts show similar features to those of modern humans (Holloway 1985) but their average cranial capacity is larger measuring approximately 1,520 cm$^3$ (from 1,200 to 1,700 cm$^3$). Large brains might be related to
cold-climate adaptation in these hominins (Churchill 1998). Even though absolute brain size was larger on average in Neanderthals relative to modern humans, their relative brain size may have been smaller due to their greater body mass (Ruff et al. 1997). The brain enlargement characteristic of both Neanderthals and modern humans appears to have followed distinct evolutionary trajectories in the two lineages, with Neanderthals retaining an archaic endocranial shape despite larger size, and modern humans exhibiting distinct shape, as well as size, changes (Bruner et al. 2003).

13.4.1.3 Basicranium

The mastoid process is small and equal in size to or smaller than the juxtamastoid eminence. It often shows a mastoid tubercle. The petrotympanic crest originates at the most inferiorly projecting part of the tympanic and the tympanic plate is coronally oriented. The mandibular fossa is wide, shallow, and medially closed-off. The foramen magnum is long, narrow, and ovoid in shape. The cranial base is relatively flattened. Recent examination of the internal morphology of the inner ear using computer tomography (CT) scans has revealed a distinctive shape for the Neanderthal bony labyrinth, most significantly characterized by an inferior placement of the posterior semicircular canal (Vallois 1969; Santa Luca 1978; Laitman and Heimbuch 1982; Trinkaus 1983; Stringer et al. 1984; Stringer 1985; Vandermeersch 1985; Hublin 1988; Condemi 1991, 1992; Elyaqtine 1995; Hublin et al. 1996; Schwartz and Tattersall 1996b; Harvati 2003b; Spoor et al. 2003).

13.4.2 Mandible

Neanderthal mandibles show a receding symphysis resulting in the absence of a mental eminence or chin. There is a space between the third molars and the ascending ramus, termed the retromolar gap. Recent analyses have shown this trait to be related to increased size in modern humans, great apes, and Middle–Late Pleistocene European fossils, suggesting caution in the interpretation of its derived status. The Neanderthal mental foramen is positioned posteriorly below the first mandibular molar rather than the premolars, unlike modern humans. The gonial area is rounded. The mandibular (sigmoid) notch is shallow and asymmetric with a coronoid process that is higher than the condyle. The mandibular notch meets the condyle in a medial position, and the condyle is laterally expanded. The submandibular and pterygoid fossae are very deep, and the mandibular foramen shows an oval-horizontal shape (Boule 1911–1913;

13.4.3 Dentition

The dimensions of the Neanderthal posterior dentition completely overlap with those of modern humans. However, Neanderthal anterior teeth, and particularly the incisors, are larger. Neanderthal teeth show enlargement of the pulp chambers (taurodontism), although this trait is variable in its degree of expression and seems to be more weakly expressed in Eastern Neanderthals. Several morphological dental features appear at very high frequencies in Neanderthals compared to modern humans. These include shoveling of the incisors and the presence of a tuberculum dentale; asymmetric lower fourth premolars with transverse crests and distolingual cusps; markedly skewed upper molars; lower molars with mid-trigonid crests; and the absence of four-cusped lower third molars. Finally, the Neanderthal dentition is distinctive in its wear patterns, showing markedly greater wear anteriorly. Although this is not a heritable trait, it has been used to infer behavioral practices (Keith 1913; Trinkaus 1983; Bytnar et al. 1994; Bailey 2002, 2004; Bailey and Lynch 2005; Bailey and Hublin in press).

13.4.4 Postcranium

The Neanderthal postcranium (Figure 13.4) is overall robust, with markedly curved shafts of the femur and radius, thick cortical bone, and strong muscle and ligament markings. Neanderthals were short relative to early modern humans and probably also to earlier H. ergaster populations. Estimated stature averages ~169 cm for males and ~160 cm for females. Body mass is estimated at ~78 kg for males and ~66 kg for females. Additional Neanderthal postcranial features include a broad and deep ribcage, with large thoracic volume, especially inferiorly; relatively short distal limb segments; large articular heads of the tibia and femur; a relatively low angle between the femoral neck and shaft; the absence of a pilaster on the femur and a more rounded (in cross section) femoral shaft; a dorsal, rather than ventral, sulcus on the axillary border of the scapula; large and round apical tufts of the manual phalanges and a relatively short proximal thumb phalanx; clavicles showing two curvatures in dorsal view; and
an elongated and thin pubic ramus (Boule 1911–1913; Trinkaus 1983, 1997; Ruff 1991, 1993, 1994; Holliday 1997a, b; Ruff et al. 1997; Churchill 1998; Rosenberg 1998; Trinkaus and Ruff 1998; Pearson 2000; Voisin 2000; Niewoehner 2001; Franciscus and Churchill 2002; Weaver 2003; Sawyer and Maley 2005). Several of these traits have been linked to high activity levels and/or cold-climate adaptation. Others could represent Neanderthal derived features, but since very little is known about their ancestral conditions, no definitive assessments can be made at present. Neanderthal body proportions are commonly viewed as “hyperarctic.” It has been suggested that the short stature and short distal limb proportions represent a cold-climate adaptation following Bergmann’s and Allen’s rules, as seen also in some modern human populations (Trinkaus 1981; Holliday 1997a, b; Steegmann et al. 2002). Overall robusticity, wide trunks, and features of the Neanderthal femur and pelvis have also been linked to climate adaptation (Ruff 1994; Pearson 2000; Weaver 2003). As improved paleoclimatic information suggests that Neanderthal ranges followed favorable climatic conditions, the designation of Neanderthals as “hyperarctic” has been challenged (Finlayson 2004). A recent estimate of the ability of the Neanderthal body shape to withstand cold temperatures showed only a small advantage over early modern humans with a less “cold-adapted” body form (Aiello and Wheeler 2003), indicating that Neanderthals could not have inhabited their high-latitude habitats without substantial cultural insulation.

### 13.5 Life history, pathology, and trauma

Neanderthal growth seems in many ways similar to modern humans, although there are indications that some aspects of their development, including brain and dental growth, may have been accelerated (Dean et al. 1986, 2001; Ramirez Rozzi and Bermudez de Castro 2004). The age mortality profile observed among Neanderthals differs from that found in recent human and other mammals in having a low percentage of older adults and infants and a high percentage of adolescents and prime age adults (Trinkaus 1995). A similar pattern has been observed in the Atapuerca Sima de los Huesos and Krapina human assemblages (Boquet-Appel and Arsuaga 1999). Although there are problems associated with such paleodemographic analyses, the observed mortality profile suggests very low adult life expectancy, probably associated with high levels of stress and trauma (Trinkaus 1995). Increased survivorship of adults resulting in a longer lifespan may have appeared very late in human evolution and not until the advent of early modern humans (Caspari and Lee 2004).
Indications of trauma and stress are ample in the Neanderthal skeletal record; so much so that it has been remarked that posttraumatic lesions can be found on almost every well-preserved adult Neanderthal skeleton (Trinkaus 1983; Berger and Trinkaus 1995; Jelinek 1994). These lesions are concentrated in the head and neck region, producing an uncommon pattern of injury that may have resulted from hunting strategies requiring proximity to large prey animals (Berger and Trinkaus 1994). Some injuries have also been argued to result from interpersonal aggression (Trinkaus 1983; Zollikofer et al. 2002). As the majority of these lesions are healed or partially healed, they have also been seen as evidence for social assistance in these hominins, as have the multiple incidents of highly worn or otherwise nonfunctional dentition (Trinkaus 1983, 1985; Lebel et al. 2001; Lebel and Trinkaus 2002).

In addition to trauma, Neanderthal remains show elevated developmental stress (Molnar and Molnar 1985; Ogilvie et al. 1989; Jelinek 1994; Berger and Trinkaus 1995). However, the degree to which this differs from stress levels in recent foraging groups is debated (Hutchinson et al. 1997; Guattelli-Steinberg et al. 2004).

### 13.6 Neanderthal genetics

Neanderthals are the first extinct human species to yield genetic information. The first glimpse of their mitochondrial DNA came with the publication of the seminal article by Krings et al. (1997). These researchers were able to recover mtDNA from the Neanderthal (Feldhofer 1) type specimen and to compare it to the homologous sequence from diverse modern human populations. The Neanderthal sequence was outside the range of modern human variation and was equally dissimilar to modern human sequences from different geographic regions. It pointed to a last common ancestor for the mitochondrial genome of Neanderthals and modern humans at approximately 500 ka (between 317 and 741 ka).

The same mitochondrial region has since been sequenced partially or in its entirety for 10 Neanderthal specimens of diverse geographic origin (Table 13.2). All have yielded similar, Neanderthal-like sequences, which group together as a distinct clade. On the other hand, all of the earliest European modern human specimens so far tested have yielded only modern humanlike and no Neanderthal-like mtDNA sequences (Caramelli et al. 2003; Serre et al. 2004). From the small number of specimens sequenced, it appears that Neanderthal mtDNA diversity was quite low and similar to that observed among modern humans today, with no evident geographic clustering (Krings et al. 2000; Schmitz et al. 2002; Lalueza-Fox et al. 2005; Serre and Pääbo in press). Thus the sequence from
the Feldhofer 2 individual is more similar to that from the Vindija 75 specimen than to Feldhofer 1 (Schmitz et al. 2002), and the Iberian El Sidrón 441 Neanderthal carries a mutation present in the Croatian specimens and Feldhofer 1 but not in the other five fossils sequenced (Beauval et al. 2005; Lalueza-Fox 2005). An estimate based on nine of these specimens placed the time to the most recent Neanderthal mtDNA ancestor to approximately 250 ka (Lalueza-Fox et al. 2005).

### Table 13.2

<table>
<thead>
<tr>
<th>Specimen</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feldhofer 1</td>
<td>Krings et al. (1997)</td>
</tr>
<tr>
<td>Feldhofer 2</td>
<td>Schmitz et al. (2002)</td>
</tr>
<tr>
<td>Mezmalskaya</td>
<td>Ovchinnikov et al. (2000)</td>
</tr>
<tr>
<td>Vindija 75</td>
<td>Krings et al. (2000)</td>
</tr>
<tr>
<td>Vindija 77</td>
<td>Serre et al. (2004)</td>
</tr>
<tr>
<td>Vindija 80</td>
<td>Serre et al. (2004)</td>
</tr>
<tr>
<td>Engis 2</td>
<td>Serre et al. (2004)</td>
</tr>
<tr>
<td>La Chapelle-aux-Saints</td>
<td>Serre et al. (2004)</td>
</tr>
<tr>
<td>El Sidrón 441</td>
<td>Lalueza-Fox et al. (2005)</td>
</tr>
<tr>
<td>Les-Roches-de-Villeneuve</td>
<td>Beauval et al. (2005)</td>
</tr>
</tbody>
</table>

13.7 Behavior

#### 13.7.1 Technology

Neanderthals are most commonly, though not exclusively, associated with the Mousterian lithic technology, named after the site of Le Moustier in the Dordogne, France. Typical of Mousterian industries was the use of both Levallois and discoidal flaking techniques for the production of flakes that could be converted to a wide range of shapes, including various kinds of side scrapers, retouched points, denticulates, notches, and sometimes small handaxes (Debénath and Dibble 1994; Mellars 1996; Shea and Brooks 2000). In addition to Europe, the Mousterian is found in the Caucasus, the Near East (where it is associated with both Neanderthals and early modern humans) and North Africa (where it is not associated with Neanderthals). Mousterian industries appear in Europe as early as ~200–150 ka and possibly earlier in the Near East, but most sites are dated to the interval from ~130 to 30 ka.

The lithic raw material used for the production of tools in most Mousterian sites tends to be locally available. Most raw materials derive from within a 5–6 km range, and only a very small component derives from distant sources (Mellars 1996).
These are mostly transported as finished tools. There is a lack of specialized use of different types of raw materials in the Mousterian, as well as a lack of specialized quarries. Very few bone tools are known. Some points appear to have been hafted and were probably used as spear points (Mellars 1996; Shea and Brooks 2000). Wooden tools were probably also made, as is evidenced by several well-preserved wooden spears discovered in Schöningen and dated to approximately 400 ka (Thieme 2000) and by parts of similar implements from Clacton-on-Sea (possibly ca. 350 ka) and Lehringen (ca. 130–110 ka; Mellars 1996).

Neanderthal sites show relatively little structure compared to later Upper Paleolithic sites. The living areas are small and exhibit no clear focus of activity. Artificial structures are rare, although exceptions are known. Hearths are well defined and were probably central in tool production and bone processing but are not consistent in their location (Mellars 1996). Controlled use of fire appears widespread in Europe from the Middle Pleistocene Oxygen Isotope Stage (OIS) 11 (~400 ka) onward, and possibly earlier (Gowlett 2006).

Until recent years, the Mousterian was commonly thought to represent a static culture. However, redating of Mousterian sites has shown changes with time in regional industries from Europe and the Near East (Shea and Brooks 2000). Additionally, reanalysis of some Mousterian sites has shown technological responses to climatic changes (Kuhn 1995; Shea and Brooks 2000). Some “transitional” Middle–Upper Paleolithic industries, like the Châtelperronian industry in France, the Uluzzian in Italy, and the Szeletian in East-Central Europe, also show strong affinities with the Mousterian. These were originally thought to have been made by early modern humans, also generally considered responsible for the Aurignacian industry. Recently, however, the Châtelperronian has been found associated with Neanderthal skeletal remains in two sites in France, St. Césaire, and Arcy-sur-Cure (Lévéque and Vandermeersch 1980; Hublin et al. 1996), both dated to approximately 35 ka, suggesting that at least some of these transitional industries were produced by late Neanderthal populations. This discovery has prompted intense debate over the identity of the makers of these industries, the possibility of Neanderthal acculturation by, or trade with, early modern humans, and the cognitive capacities and ability for symbolic thought in Neanderthals (D’Errico et al. 1998; Zilhão and d’Errico 1999; Mellars 1999, 2005; Klein 2000; Harvati et al. 2003; Bar-Yosef 2005; Gravina et al. 2005; Svoboda 2005).

13.7.2 Subsistence

Neanderthal sites abound in faunal remains of various taxa, indicating a high reliance on meat in their diet. Large palaearctic mammals are most commonly found in these assemblages, including bison, wild cattle, horse, reindeer, red and
fallow deer, ibex, wild boar, and gazelle (Shea and Brooks 2000). Sites from the high latitudes of Northern and Central Europe indicate an almost exclusive reliance on large- to medium-sized mammals, with very little small game and low diversity of animals consumed (e.g., very few bird or fish remains; Hockett and Haws 2005). Middle-latitude Neanderthal sites (Southwestern France and Northern Spain) also show low diversity and a focus on terrestrial mammals but indicate a somewhat greater reliance on medium-sized mammals. Sites from the Mediterranean region still show reliance on large- and medium-sized terrestrial mammals, but also preserve evidence for consumption of other food sources, such as shellfish, birds, and marine mammals (Stiner 1994; Barton 2000; Currant 2000; Hockett and Haws 2005). Plant remains in Neanderthal sites are quite rare, likely due to their poor preservation. Phytoliths and other vegetal remains known primarily from Mediterranean sites point to a plant component in Neanderthal diet which probably included wild legumes and grasses as well as seeds and fruit (Gale and Carruthers 2000; Madella et al. 2002; Lev et al. 2005).

In addition to direct evidence of faunal and plant remains from archeological sites, Neanderthal diets can be assessed using the isotopic signature of the Neanderthal skeletons themselves. Analysis of the stable isotopes of carbon and nitrogen has now been undertaken for a number of Neanderthal specimens from a wide time range (~130–30 ka) and so far has invariably indicated a very strong reliance on herbivore meat (Fizet et al. 1995; Bocherens et al. 1999, 2005; Richards et al. 2000). All Neanderthal bones so far analyzed are similar to top predators in their isotopic composition. Furthermore, some of the isotopic studies suggest a much greater reliance on very large herbivores, such as wholly rhinoceros or wholly mammoth, than had been previously thought based on the faunal archeological evidence (Bocherens et al. 2005). The isotopic analyses agree with zooarcheological studies in suggesting a very small component of marine foods in Neanderthal diets, in sharp contrast with later, Upper Paleolithic modern humans (Richards et al. 2001, 2005). However, no isotopic analysis has yet been conducted for Neanderthal specimens from the Mediterranean coast, whose diets may have been more varied.

The degree to which Neanderthals obtained meat through hunting as opposed to scavenging has been a subject of debate (Binford 1983; Chase 1986; Stiner 1990). Some have suggested that the relative significance of these two activities probably varied seasonally and from region to region (Shea and Brooks 2000). Among the arguments brought forth to support scavenging as the primary Neanderthal activity is the high proportion of cranial faunal remains in Neanderthal sites and the age-profile of the remains, thought to be for the most part old individuals rather than prime-age adults (Binford 1983). However, a bias toward cranial remains may simply represent a bias in the butchering and transportation
of large carcasses (Mellars 1996). The age-mortality profile is now known to vary, with several French Middle Paleolithic sites showing a catastrophic mortality profile inconsistent with hypotheses of scavenging (summarized in Mellars 1996). Further evidence in support of hunting comes from the wooden spears from Schöningen dated to 400 ka (Thieme 2000), although there is disagreement as to whether these represent throwing or thrusting spears. Finally, the Neanderthal stable isotopic signature suggests active predation on the part of the Neanderthals and is difficult to reconcile with a subsistence strategy consisting primarily of scavenging (Richards et al. 2000).

### 13.7.3 Symbolic thought and language

The Neanderthal ability for symbolic thought and language is hotly debated. The archeological record shows a dearth of “symbolic” objects, such as objects of art or personal ornamentation, in Mousterian assemblages, compared not only with later Upper Paleolithic industries (Mellars 1996) but also with some penecontemporaneous African sites (McBrearty and Brooks 2000; Henshilwood et al. 2001). The lack of such objects is often argued to indicate a lack of human cognitive abilities and language. However, it has also been pointed out that the archeological record is a very limited and imperfect record of behavior, perhaps in this case resulting in a biased documentation (or lack thereof) of Neanderthal symbolic activities. The recent discovery of Neanderthal remains associated with transitional industries, and, in the case of Arcy-sur-Cure, with lavish personal ornaments, has raised tremendous discussion over the identity of the makers of these objects, as well as the processes that would have led to their association with Neanderthals (i.e., trade, acculturation, or endogenous development; see e.g., Hublin et al. 1996; D’Errico et al. 1998; Zilhão and d’Errico 1999; Mellars 1999, 2005; Klein 2000; Harvati et al. 2003; Bar-Yosef 2005).

Evidence in support for Neanderthal ability for some symbolic thought is the occurrence of ochre and manganese “crayons” in Neanderthal sites and the burial of at least some Neanderthal skeletal remains. Although the apparent Neanderthal burials have been argued to be simply the product of natural processes (Gargett 1999), the recovery of a number of largely complete skeletons from diverse sites found in articulation and placed in shallow pits is strongly indicative of intentional burial. Nevertheless, evidence for grave goods and other burial practices is scant and controversial (Mellars 1996; Shea and Brooks 2000).

In terms of the anatomical evidence for language and cognition, Neanderthals possessed, cranial capacities as large as or larger than modern humans. Their endocasts show similar features to those of modern humans and similar
left–right asymmetries (Holloway 1985) although they retain an “archaic” overall shape (Bruner et al 2004). The relatively flat Neanderthal cranial base was long considered to indicate a larynx positioned so high that it would preclude the production of certain speech sounds and particularly of vowels crucial to speech perception (Laitman and Heimbuch 1982; Lieberman 1989). However, this assessment was based on a male model for the modern human vocal tract, which shows secondary growth in laryngeal position for reasons unrelated to speech production. When a female model of the human vocal tract is used, the Neanderthal laryngeal position falls within the modern human range (Clegg 2004). The size of the hypoglossal canal, which transmits the nerves to the exceptionally large human tongue musculature, is similar in Neanderthals and modern humans and larger than in earlier hominins (Kay et al. 1998). This evidence suggests a similar function for the Neanderthal tongue in speech production, but its significance has been questioned (DeGusta et al. 1999). Neanderthals are also similar to modern humans and unlike earlier hominins in their enlarged thoracic vertebral canals, which could indicate an expansion of thoracic innervation (MacLarnon and Hewitt 1999). The resulting greater control of the intercostal musculature would enhance breathing control and could indicate the ability for speech. Finally, the anatomy of the outer and middle ear in the Middle Pleistocene pre‐Neanderthal fossils from Sima de los Huesos (Atapuerca, Spain) was found to be similar to that of modern humans and specialized for speech perception (Martínez et al. 2004), supporting speech capabilities for Neanderthals and their ancestors.

13.8 Evolution and classification

13.8.1 The “Accretion Hypothesis”

The “Accretion Model” for the evolution of Neanderthals (Dean et al. 1998; Hublin 1998) accounts for the progressive appearance of Neanderthal morphology through time, beginning around 450 ka (OIS 12). According to this hypothesis, the Neanderthal lineage became isolated in Europe due to the severe climatic conditions of the Pleistocene. The geographic range of Mid‐Pleistocene European hominins would have been restricted during glacial maxima by ice sheets covering northwestern Europe and by their associated permafrost zones. Furthermore, formidable barriers would impede any movement: to the South the Mediterranean Sea, to the East the mountain chains of the Caucasus and their glaciers, as well as an expanded Caspian Sea (Hublin 1998). In these conditions of isolation, the Neanderthal morphology is thought to have become gradually fixed, partly
through natural selection as an adaptation to cold-climate conditions but perhaps primarily through the process of genetic drift.

Although the accretion process of Neanderthal features is mosaic in nature, facial and mandibular features become established first, followed by features in the occipital region and finally in the temporal bone and vault (Dean et al. 1998; Hublin 1998; Rightmire 1998). Four broad stages of Neanderthal evolution have been described in this scenario (Dean et al. 1998; Hublin 1998). Stage 1 includes “early pre-Neanderthals,” i.e., the Middle Pleistocene archaic specimens, such as Petralona, Arago, and Mauer, dating from before OIS 12. These hominins are considered to show incipient Neanderthal features mainly in the facial region. Stage 2 (OIS 11-9) specimens are termed “pre-Neanderthals” (e.g., Steinheim, Swanscombe). They are thought to exhibit Neanderthal morphology more clearly, showing Neanderthal features also in the occipital area. Stage 3 (OIS 7-5, Biache, Krapina, Saccopastore) “early Neanderthal” specimens show most Neanderthal traits in the posterior cranium and some also in the temporal region. Finally, Stage 4 comprises the “classic Neanderthals” of OIS 4 and 3 (Neanderthal, La Chapelle-aux-Saints, Amud), showing fully expressed Neanderthal morphology.

13.8.2 Classification of Middle and Late Pleistocene humans

According to the accretion hypothesis, Neanderthal evolution was an anagenetic process with no speciation event resulting in the appearance this taxon. Within the framework of this model, Neanderthals can be viewed either as a subspecies of *H. sapiens* or as a full, distinct species, *H. neanderthalensis*. If the latter classification is accepted, then the position of the Middle Pleistocene specimens from Europe must be clarified. Traditionally, these have been included in the species *H. heidelbergensis* in which African Middle Pleistocene humans have also been placed. This taxon is viewed by some as ancestral to both Neanderthals and modern humans (Rightmire 1998). Since it seems clear, however, that the European *H. heidelbergensis* was ancestral to Neanderthals, it has also been argued that this sample should be placed within the Neanderthal lineage and within the taxon *H. neanderthalensis* (Stringer 1995; Arsuaga et al. 1997). Alternatively, the European lineage could be arbitrarily split into two paleospecies, the earlier segment retaining the nomen *H. heidelbergensis* and the later *H. neanderthalensis*. In either case, the African Middle Pleistocene specimens would have to be placed into another taxon, possibly *H. rhodesiensis* or *H. helmei* (Stringer 1995).

An alternative interpretation (Rightmire 1998; Lalueza-Fox et al. 2005; Rosas et al. in press) is that, after Middle Pleistocene populations became isolated in
Europe, there was a speciation event leading to the evolution of Neanderthals. If this were the case, then the taxon *H. neanderthalensis* should include only the specimens postdating this event, while *H. heidelbergensis* can include both African and European Middle Pleistocene human fossils (Rightmire 1998). Some researchers consider the early skeletal remains from Gran Dolina Atapuerca (dated to \(~780\) ka) as a distinct species, *H. antecessor*, itself the last common ancestor of both Neanderthals and modern humans (Bermudez de Castro et al. 1997). As these fossils are fragmentary and mostly subadult, however, their phylogenetic position and relationships cannot be determined with certainty at present.

13.8.3 The Neanderthal role in modern human evolution

Ever since their assignment to a distinct species, *Homo neanderthalensis*, by King (1864), the taxonomic position of Neanderthals and their role in human evolution have been the subject of intense debate. It is now commonly recognized that Neanderthals and earlier Middle Pleistocene European fossils form a distinct evolutionary lineage, at least partly geographically isolated in Western Eurasia, and sharing a common ancestor with modern humans in the early Middle Pleistocene. There is no agreement, however, as to whether the last representatives of this lineage, the “classic” Neanderthals, were *reproductively* isolated from early modern humans arriving in Europe some 40 ka, and as to how much, if at all, they were able to contribute to the gene pool of modern people.

The question of Neanderthal species status is partly dependent on the species concept used. Two commonly used species concepts in paleontology are the phylogenetic (PSC) and the evolutionary species concepts (ESC). Under both of these definitions, Neanderthals can be considered distinct species. The unique, derived status of a suite of Neanderthal features has been recognized for a number of years, pointing to distinct species status under the PSC (Hublin 1978; Santa Luca 1978; Stringer et al. 1984; Tattersall 1986, 1992, 2000; Stringer and Andrews 1988; Schwartz and Tattersall 1996a, b). Furthermore, these distinctive traits appear early in ontogeny, possibly in the fetal stage, and there are some indications that Neanderthal development differed in its pace from that of modern humans (Dean et al. 1986; Smith 1991; Maureille and Bar 1999; Ponce de León and Zollikofer 2001; Krovitz 2003; Ramírez Rozzi and Bermúdez de Castro 2004; Zollikofer and Ponce de León in press). Neanderthals and their Middle Pleistocene European predecessors are commonly thought of as a distinct evolutionary lineage (Dean et al. 1998; Hublin 1998), fitting the definition of a species under
the ESC. In terms of magnitude of morphological difference in the context of primate patterns of variation, a recent study of 11 primate species and their subspecies showed that the difference between Neanderthals and modern humans, including Upper Paleolithic Europeans, is much larger than the differences between cospecific subspecies, and as large or larger than the differences between cogeneic species (Harvati 2003a; Harvati et al. 2004). This evidence indicates that species status is most appropriate for primate taxa showing this level of morphological differentiation. Distinct species status for modern humans relative to Neanderthals and other “archaic” hominins has also been argued on the basis of the uniquely derived patterns of cranial architecture in modern humans and the retention of the archaic pattern in Neanderthals (Lieberman et al. 2002).

It is more difficult to use the biological species concept (BSC), often the preferred definition of species. As it uses reproductive isolation for its criterion for species delineation, it cannot be applied directly to the fossil record. Nevertheless, some paleoanthropologists consider that Neanderthals and modern humans interbred successfully and that the former are at least partially ancestral to Upper Paleolithic and modern Europeans. They see evidence for Neanderthal-like features in early modern European specimens and for trends of “modernization” in some late Neanderthal samples (Smith 1982, 1992; Frayer et al. 1993; Ahern et al. 2002; Trinkaus et al. 2003). However, others find no evidence of intermediate morphology (Bräuer and Broeg 1998; Harvati 2003a; Bräuer et al. 2004, in press; Harvati et al. 2004). A recent claim for a Neanderthal–modern human hybrid from Portugal dated to approximately 24 ka (Duarte et al. 1999) is difficult to substantiate due to the subadult status of the specimen and to its relatively late geological age (Tattersall and Schwartz 1999). From a molecular perspective, the mtDNA of Neanderthals and Upper Paleolithic Europeans shows no evidence for admixture between the two groups (Serre et al. 2004). Recent demographic models of the Neanderthal–modern human interaction based on the ancient DNA available data suggest negligible contribution of Neanderthals to the modern human gene pool (<1–2%; Currat and Excoffier 2004; Weaver and Roseman 2005). Therefore, even though interbreeding might have been possible, it seems not to have occurred to any level of significance for modern human evolution.

13.9 The Neanderthal extinction

The last appearance date of Neanderthals is ca. 30 ka, with modern humans appearing in Europe for the first time approximately 40 ka. Although there are problems with radiocarbon dating due to modern carbon contamination and
to the lack of accuracy of the method during this time interval, it appears that Neanderthals and early modern humans probably overlapped in Europe for several millennia (Stringer et al. 2003). For some researchers, the last appearance of Neanderthals simply signifies their evolution into modern Europeans (Brace 1995) or their assimilation into the modern human gene pool (Smith et al. 2005; Svoboda 2005). Others, however, view it as a true extinction, even if limited genetic exchanges with modern humans took place.

Several scenarios for the Neanderthal extinction have been proposed, and often they invoke some direct or indirect competition with early modern humans arriving in Europe shortly before Neanderthals disappear. Proposed modern human competitive advantages range from demographic factors, such as small differences in birth and mortality rates or in interbirth spacing (Zubrow 1989; Skinner 1997; Flores 1998), to larger group sizes (Gat 1999), to greater dietary diversity (Richards et al. 2001; Hockett and Haws 2005). The role of climate was until recently not emphasized, as Neanderthals disappeared in OIS 3, during what was thought to be relatively stable conditions preceding the last glacial maximum (Stringer et al. 2003). Some recent hypotheses, however, consider climatic and environmental factors to be major driving forces. It has been proposed that habitat degradation and fragmentation led to the disappearance of Neanderthal populations, with modern humans arriving in areas previously occupied by Neanderthals after the latter were already extinct (Finlayson 2004). The Neanderthal demise has also been viewed as one of the many Late Pleistocene megafauna extinctions caused by the loss of an environment with no modern analogue (Stewart et al. 2003; Stewart 2005).

Recently available detailed paleoclimatic records have shown that OIS 3 was dominated by much more unstable climatic conditions than previously thought, leading to a reexamination of the role of climate deterioration in the Neanderthal demise (van Andel and Davies 2003). Modeling of climatic stress (defined as the indirect effects of environmental change) based on these new data found two stress peaks at \( \sim 65 \) and \( 30 \) ka, the second one appearing to be more prolonged and severe than the first one and possibly directly related to the Neanderthal extinction (Stringer et al. 2003). However, as Neanderthals had survived previous cold phases, it is difficult to accept climate change as the sole reason for their demise. The argument has been made that no two cooling episodes are the same and that taxa surviving one cooling episode may not persist through the next (Stewart 2005). An alternative interpretation, however, sees the advent of modern humans, perhaps with better cultural buffeting and more effective social networks, as providing the \textit{coup de grâce} to the highly stressed Neanderthal populations through competition for severely limited resources during the critical time
interval of their coexistence in Europe (Stringer et al. 2003; see also Lahr and Foley 2003). In this view, it is the interaction between the effects of fluctuating climate and environment and of competition with modern humans that led to the eventual Neanderthal extinction.

References


Cahiers de Paléoanthropologie. CNRS éditions, Paris

Gargett RH (1999) Middle Palaeolithic burial is not a dead issue: The view from Qafzeh, Saint-Césaire, Kebara, Amud, and Dederiyeh. J Hum Evol 37; 27–90


processes of Western Europe. CR Acad Sci Paris IIa 321: 931–937
perturbation in the Krapina Neandertals. Curr Anthropol 38:904–914
Mandibular condyle traits in Neandertals and other Homo: A comparative,correlative,
Jelinek A(1994)Hominids,energy,environment
and behavior in the Late Pleistocene. In:Nitecki MH,Nitecki DV(eds)Origins of
Kay RF,Cartmill M,Balow M(1998)The
hypoglossal canal and the origin of human
vocal behavior. Proc Natl Acad Sci USA 95:
5417–5419
Keith A(1913)Problems relating to the earlier
forms of prehistoric man. Proc R Soc Med
(Odontol)6:103–119
King W(1864)The reputed fossil man of the
Neanderthal. Q J Sci 1:88–97
Klein RG(2000)Archaeology and the evolu-
tion of human behavior. Evol Anthropol 9:
7–36
Klein RG(2003)Whither the Neandertals?
Science 299:1525–1527
Krings M,Stone A, Schmitz RW,Krainitzki H,
DNA sequences and the origin of modern
humans. Cell 90:19–30
Krings M,Capelli C,Tsentscher F,Geisert H,
Meyer S,von Haessler A,Crossschmidt K,
view of Neandertal genetic diversity. Nat
Genet 26:144–146
Krovitz GE(2003)Shape and growth differ-
ences between Neandertals and modern
humans grounds for a species-level distinc-
tion. In: Thompson J,Krovitz G,Nelson A
(eds)Patterns of Growth and Development
in the Genus Homo. Cambridge University
Press, Cambridge, pp 320–342
Kuhn SL(1995)Mousterian Lithic technology:
an ecological perspective. Princeton University
Press, Princeton
human diversity: A study of cranial variation.
Cambridge University Press,Cambridge
Lahr MM,Foley RA(2003)Demography, dis-
persal and human evolution in the Last Glacial Period. In: van Andel T,Davies W
(eds)Neandertals and modern humans
in the European landscape during the last
 glaciation. McDonald Institute for Archaeo-
logical Research,Cambridge,pp 241–256
Laitman JT,Heimbuch RC(1982)The basicra-
nium of Plio-Pleistocene hominids as an
indicator of their upper respiratory systems.
Am J Phys Anthropol 59:323–344
Lalueza-Fox C,Sampietro ML,Caramelli D,
Puder Y,Lari M,Calafell F,Martinez-Maza
C,Bastir M,Fortea J,Rasilla M,Bertranpetit
J,Rosas A(2005)Neandertal evolutionary
genetics, mitochondrial DNA data from the
Iberian Peninsula. Mol Biol Evol 22:
1077–1081
human remains from the Bau de l’Aubesier.
J Hum Evol 43:659–685
Lebel S,Trinkaus E,Faure M,Fernandez P,
Guérin C,Richter D,Mercier N,Valladas H,
Wagner GA(2001)Comparative morpholo-
gy and paleobiology of Middle Pleistocene
human remains from the Bau de l’Aubesier,
Vaucluse,France. Proc Natl Acad Sci USA 98:
11097–11102
Lev E,Kislev ME,Bar-Yosef O(2005)Mouste-
rian vegetal food in Kebara Cave, Mt.
derestes humains dans un niveau castelperronien à Saint-Césaire (Charente-Maritime).
about recent human evolution from skulls:
Integrating morphology, function, develop-
ment and phylogeny. Curr Anthropol 36:
159–197
Lieberman DE,McBratney BM,Krovitz G
(2002)The evolution and development of
human form in Homo sapiens. Proc Natl
Acad Sci USA 99:1134–1139


Rightmire GP (2001) Patterns of hominid evolution and dispersal in the Middle Pleistocene. Quat Int 75: 77–84
Institute for Archaeological Research, Cambridge, pp 233–240
Trinkaus E (2003) Neandertal faces were not long; modern human faces are short. Proc Natl Acad Sci USA 100: 8142–8145
Weaver TD (2003) The shape of the Neandertal femur is primarily the consequence of a hyperpolar body form. Proc Natl Acad Sci USA 100: 6926–6929
14 Origin of Modern Humans

Günter Bräuer

Abstract

In the early 1980s, a new period in the debate on modern human origins began, focusing on two alternatives, the Multiregional Evolution model and the Out-of-Africa hypothesis. Over the last decades, new hominid discoveries, absolute dating, and other evidence have supported the latter view which proposes a recent common origin of modern humans in Africa. The increasing evidence made the idea of long-term regional evolution up to modern humans in Europe and Asia, following the first expansion out of Africa at nearly 2 Ma, more and more unlikely. Only the African fossil record documents a continuous early modernization process. In contrast, the European evidence shows a replacement of the Neanderthals by modern humans. Also, most of the claimed evidence for regional continuity in China and Australasia has turned out to be unsubstantiated. Major questions of the current discussion on modern human origins refer to the evidence for gene flow during the replacement period and to the number of species involved in Middle Pleistocene evolution.

14.1 Early and current controversies

The origin of modern humans has always been a controversial topic in paleoanthropology. Already at the beginning of the twentieth century and based on a very sparse—mainly European—fossil record, alternative phylogenetic scenarios were suggested. Schwalbe (1906) proposed a unilinear concept in which Neanderthals were an intermediate form between Dubois’ *Pithecanthropus* and modern humans. Boule (1913), on the other hand, saw the Neanderthals as an evolutionary cul-de-sac postulating a parallel “presapiens” lineage to modern humans. Although the Presapiens hypothesis became widely accepted in the subsequent decades, some researchers, especially Hrdlička (1927) and Weidenreich (1943) supported unilinear concepts. Weidenreich (1943), after examining the newly discovered Chinese and Javanese hominid fossils, proposed his Polycentric Evolution theory which suggested different evolutionary lines in North Asia, Southeast Asia/Australasia, Europe/Near East, and eastern/southern Africa. He regarded the regional sequences as an interconnected web evolving in a single
common direction. Yet, Weidenreich’s (1947) explanation of orthogenesis as
the driving factor of such a polycentric evolution was rejected by the new syn-
thetic theory of evolution. In addition, Coon’s (1962) later model of largely
isolated parallel evolution in different parts of the world was eventually dis-
missed. Thus, according to Wolpoff et al. (1984 p 418) during the 1960s and
1970s, only a few researchers favored local continuity between modern and
ancient populations.

In the 1950s, new support emerged for the Presapiens hypothesis, in partic-
ular by the cranial remains from Fontéchevade in connection with the specimens
from Swanscombe and Steinheim (Valloir 1954). However, later research on the
critical presapiens specimens revealed their affinities to Neanderthals and
Preneanderthals (Trinkaus 1981; Hublin 1982). In addition, new hominin dis-
covers, such as the partial crania from Arago and Biache St. Vaast, demonstrated
that the idea of a separate lineage to modern humans in Europe was no longer
tenable. It became clear that there was only one lineage in Europe leading to
Neanderthals (Bräuer 1984a). During the 1950s another model, the Prenean-
derthal hypothesis, was proposed (Howell 1951) which assumed that the lineages
to Neanderthals and modern humans only split during the Eem Interglacial.
Following this event, the southwest Asian “progressive” Preneanderthals evolved
into modern humans, while the European Preneanderthals developed into the
robust “classic” Neanderthals. However, this concept was also not supported by
later research. No diachronic trends of reduction in size could be observed in the
Near Eastern Neanderthals, and later dating revision revealed that early modern
humans and Neanderthals were nearly contemporaneous in this region. Thus, by
the late 1970s, the question of the origin of modern humans was again largely
open. A few researchers such as Wolpoff (1980) continued to favor evolutionary
continuity in Europe and elsewhere, whereas others like Howells (1976) assumed
a recent common origin of modern humans.

In the early 1980s, a new period in the controversy on modern human origins
began, mainly focusing on two alternatives, the Multiregional Evolution model
and the Out-of-Africa hypothesis (Figure 14.1). Wolpoff et al. (1984) proposed
the Multiregional Evolution model, which was largely based on Weidenreich’s
theory of Polycentric Evolution. In contrast to Weidenreich’s (1947) explanation
of orthogenesis, the supporters of the new multiregional concept see a balance of
evolutionary forces, such as gene flow, local selection, and drift, as explanations
for the assumed long-term regional continuities (Wolpoff 1992). Yet they accept-
ed and used most of Weidenreich’s morphological observations for regional
continuity that had been collected in the 1930s and 1940s (Wolpoff et al. 1984;
Frayer et al. 1993). Although assuming interregional gene flow, multiregionalists
have emphasized the regional pattern of evolution as stated, for example, by
### Figure 14.1
(a) Multiregional evolution (after Thorne and Wolpoff 2003 p 52) and (b) Out-of-Africa models

<table>
<thead>
<tr>
<th>Europe and Levant</th>
<th>Africa</th>
<th>East Asia</th>
<th>Australasia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lagar Velho</td>
<td>Aflou</td>
<td>Shandong</td>
<td>Kow Swamp</td>
</tr>
<tr>
<td>Předmosti</td>
<td>Lukenga</td>
<td>Ziyang</td>
<td>Keller</td>
</tr>
<tr>
<td>Miadeč</td>
<td></td>
<td>Liajiang</td>
<td>Willandra</td>
</tr>
<tr>
<td>Vindija</td>
<td></td>
<td></td>
<td>Lakes 50</td>
</tr>
<tr>
<td>Kebara</td>
<td>Dades Soltan</td>
<td>Maba</td>
<td>Lake Mungo 1, 3</td>
</tr>
<tr>
<td>La Ferrassie</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>La Chapelle</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Qafzeh</td>
<td>Klies</td>
<td>Dingsun</td>
<td></td>
</tr>
<tr>
<td>Krapina</td>
<td>Omo Kibish</td>
<td>Xujia 20</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ehringsdorf</td>
<td>Ngaloba</td>
<td>Dali</td>
<td>Sambungmachan 1, 3</td>
</tr>
<tr>
<td>Blache</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zuttiye</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sima de los Huesos</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Petralona</td>
<td>Kabwe</td>
<td>Zhokoudian H</td>
<td></td>
</tr>
<tr>
<td>Arago</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Steinheim</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gran Dolina</td>
<td>Bodo</td>
<td>Zhokoudian D, E, L</td>
<td>Sangiran 2, 10, 12, 17</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Trinil</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Konso Gardula</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lake Turkana</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(east) 992</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dmanisi</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lake Turkana</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(east) 730, 3883, 3733</td>
<td></td>
<td>Sangiran 4, 27, 31</td>
<td></td>
</tr>
<tr>
<td>(west) 15000</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**a**
![Multiregional evolution diagram](image)

**b**
![Out-of-Africa models diagram](image)
Frayer et al. (1993 p 41): “Each geographic region we examined contains a wealth of information that shows the continuous evolution of Homo populations over time. We find neither specimens nor traits that could reflect an infusion of any African genes and their so-called more-modern morphology.” Frayer (1992 p 49) considered the “Neandertals as the probable ancestors of the people in the Upper Palaeolithic,” and Thorne (1993 p 173) is convinced that the “descendants of the Java and Peking people do not become extinct but give rise, without African influence, to the modern people of their region.”

The roots of the alternative Out-of-Africa model can be traced back to the early 1970s. Although at that time practically all evidence from archeology and paleoanthropology pointed to the presence of archaic hominids—the so-called “Rhodesioids” (named after the Kabwe cranium from Zambia)—in eastern and southern Africa from only 30 or 40 Ka (Clark 1970) new skeletal remains from Omo Kibish, Ethiopia, and new dates for the South African Border Cave specimens, yielded some indications of the early presence of modern humans at about 100 Kyr BB, or even slightly earlier (Leakey et al. 1969; Protsch 1975). Yet it was puzzling how such early moderns fit in with the much later presence of the archaic humans. Further research on the dating of the African Stone Age during the 1970s led to a drastic extension in time of the Middle and Later Stone Age, and thus to older dates of the associated hominids (Clark 1979). It also turned out that the archaic “Rhodesioids” were considerably older than had been thought.

In the late 1970s, I started a new analysis of the Middle and Late Pleistocene hominid material from Africa. This research provided a new framework of Homo sapiens evolution, suggesting a mosaic-like, continuous anatomical process of modernization leading to an early emergence of modern humans. Based on this framework and a review of the fossil evidence from Europe and the Far East, I proposed an Out-of-Africa model initially dubbed the “Afro-European sapiens” hypothesis (Bräuer 1982) because the best evidence for replacement came from Europe. Yet the model was a global one, regarding replacement as the most likely process for the Far East as well (Bräuer 1984b). The Out-of-Africa model suggested an evolution to modern humans only in Africa and subsequent dispersals into Asia and Europe that replaced the resident archaic populations, including the Neanderthals. “Replacement” was assumed to also allow for interbreeding (Bräuer 1984c p 395). A few years later, Cann et al. (1987) provided important support for a recent common origin of modern humans in Africa based on the human mitochondrial DNA (mtDNA). Although no strongly divergent mtDNA lineages could be found among extant humans, Cann (1992 p 71) did not assume reproductive isolation between the archaic and the
dispersing modern populations but rather concluded that we are about 30 Kyr too late to see the persistence of Neanderthal maternal lineages.

These and additional results from molecular biology and paleoanthropology during the 1980s supporting a recent African origin did not lead to the dismissal of the alternative Multiregional Evolution model but instead to a heated debate and an artificial polarization. The multiregionalists now claimed that the mtDNA results must be interpreted as excluding any gene flow and that this is an essential assumption of the Out-of-Africa model (Wolpoff and Thorne 1991). Although this view was rejected (Bräuer 1989, 1992; Stringer 1992), the multiregionalists focused on the criticism of the extreme “Eve Theory,” assuming that by excluding gene flow, any possible indication of regional continuity outside Africa would be sufficient to disprove the Out-of-Africa model (Frayer et al. 1993). This argument, however, was misleading, as the “Eve” concept is based on a particular interpretation of the mtDNA data and cannot be equated with the Out-of-Africa model, which includes the evidence from fossils and nuclear DNA as well (Stringer and Bräuer 1994 p 416). Thus, the Out-of-Africa replacement view allows for the possibility of gene flow between archaic and modern humans and is ready to accept any convincing evidence for it that might be found in the fossil record (Bräuer and Stringer 1997). For example, Serre et al. (2004) searched for Neanderthal mtDNA in a number of early modern specimens from Europe and emphasized that a much larger sample of early moderns would have to be studied to exclude a Neanderthal mtDNA contribution. Also, with regard to the nuclear DNA, Gibbons (2001 p 1052) stated that no one can rule out the possibility that some of us could have inherited nuclear DNA from Neanderthal or *H. erectus* and that there is no clear genetic test to detect such archaic DNA. Consequently, from the genetic point of view, it not only makes little sense to claim that the Out-of-Africa model excludes interbreeding but it is also meaningless to attempt to disprove the “Eve Theory” in order to reject the Out-of-Africa hypothesis. Currently, most supporters of the Out-of-Africa model see evidence for only a small amount of gene flow (Bräuer and Broeg 1998; Bräuer 2001a, 2006; Stringer 2001a) whereas others are even skeptical regarding any such indications of interbreeding (Tattersall and Schwartz 2000; Tattersall 2003). In addition, some previous supporters of the Multiregional Evolution perspective have meanwhile arrived at a view very close to that of the Out-of-Africa model by accepting both an African origin of modern morphology and its introduction into Europe by population movements, but assuming a more significant assimilation of the Neanderthals (Churchill and Smith 2000). In the following paragraphs, the current evidence for regional continuity and discontinuity will be considered for the different parts of the Old World.
14.2 African emergence of modern humans

Over the last two decades, the framework of the anatomical modernization process in Africa as suggested in the early 1980s (Bräuer 1982, 1984c) has been basically supported (Klein 1999; Bräuer 2001b; Stringer 2002a; Mbu and Bräuer in press). Revisions of the framework mainly concerned the chronology of the process due to new dating evidence for several hominid specimens (Clark et al. 1994; Bräuer et al. 1997; McDougall et al. 2005). The current evolutionary scheme (Figure 14.2) appears rather well founded on quite a number of diagnostic as well as absolutely dated specimens that support a mosaic-like, continuously evolving, lineage from archaic to modern humans.

The anatomical modernization process can be divided into three grades of *H. sapiens*, each of which includes hominid specimens of a similar evolutionary level (Bräuer 1989, 2001b). The specimens grouped within the early archaic *H. sapiens* category are derived relative to *H. erectus*, especially regarding their enlarged cranial capacity, more vertically oriented lateral walls, expanded frontal bone, less strongly angulated occipital bone, more vertically oriented upper scale of the occipital, higher temporal squama, and reduced development of the supraorbital and occipital tori. The late archaic *H. sapiens* is clearly more derived compared to the morphological pattern of the early archaics, especially evident in the large cranial capacity, the more reduced supraorbital torus, and the near-modern or modern face, including the canine fossa and inframalar incurvature. This grade of evolution is followed by anatomically modern *H. sapiens* with a fully modern morphology of vault and face. There is obvious continuity between the grades, which are best seen as a way of describing the levels of the modernization process. In this respect, the late archaics have also been designated as the transitional group (Smith 2002). This approach does suggest neither anything about underlying factors of anatomical modernization nor whether there are parts of the lineage that show more relevant changes toward the modern morphology than others. Thus, the identification of major structural elements in this process, such as neurocranial globularity and facial retraction (Lieberman et al. 2002), is relevant. Indeed, many of the features and aspects characterizing the grades are connected with such general changes.

A key specimen of early archaic *H. sapiens* is the Bodo hominid from Ethiopia, dated by Ar/Ar to ca. 600 Kyr BP (Clark et al. 1994). The large cranial capacity of nearly 1,300 cm³ is associated with some parietal bossing, a coronally expanded frontal, and derived features of the temporal. The supraorbital torus even shows some division into a medial and lateral portion (see also Volume 3 Chapter 12). Still, the massive face looks rather archaic. Another specimen of similar or slightly younger age (Klein 1999) is the Saldanha (or Elandsfontein) cranium from South Africa. Having an estimated cranial capacity of around 1,225 cm³, the
Figure 14.2
Fossil record of *Homo sapiens* evolution in Africa

<table>
<thead>
<tr>
<th>North</th>
<th>East</th>
<th>South</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taramsa</td>
<td>Die Kelders Cave</td>
<td>Present (ka)</td>
</tr>
<tr>
<td>Dar-es-Soltane</td>
<td>Border Cave</td>
<td>Anatomically Modern</td>
</tr>
<tr>
<td>Témara</td>
<td></td>
<td><em>Homo sapiens</em></td>
</tr>
<tr>
<td>Zouhra Cave</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haua Fteah</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Herto</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Omo 1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Omo 2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Elie Springs</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wadi Dagadélé</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Laetoli 18</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ileret / ER-3884</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ileret / ER-999</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Florisbad</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cave of Hearths</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Kabwe</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Olduvai H11</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Eyasi</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sidi Abderrahman</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Salé</td>
<td></td>
</tr>
<tr>
<td>Thomas Quarry</td>
<td>Saldanha</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bodo</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tighenif</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Late Archaic <em>Homo sapiens</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Early Archaic <em>Homo sapiens</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Developed <em>Homo erectus</em></td>
</tr>
</tbody>
</table>
Parietals are well arched and show some bossing. Also, the frontal squama is coronally enlarged. The occipital is less angulated than generally seen in H. erectus and the transverse torus is reduced. The other well-preserved early archaic specimens, such as the crania from Kabwe, Zambia, Eyasi and Ndufu, Tanzania, and Salé, Morocco, date from the same time period between 600 and 300 Ka and exhibit similar derived sapiens-like conditions (Bräuer 2001b, 2004; Mbu a and Bräuer in press).

The late archaic grade also comprises specimens spreading from northern to southern Africa. A good example of these near-moderns is the cranium KNM-ER 3884 from Ileret, East Turkana (Figure 14.3) directly dated by gamma-ray spectrometry to ca. 270 Kyr BP (Bräuer et al. 1997). A previous analysis showed that most of the cranial vault falls close to the range of Holocene Africans (Bräuer et al. 1992a). However, the cranium also exhibits a continuous supraorbital torus that deviates from the generally rather modern impression of the specimen (Bräuer 2001b; Schwartz and Tattersall 2003). Further support for such an early presence

Figure 14.3
Late archaic and early modern Homo sapiens: (a) KNM-ER 3884/ Ileret; (b) Omo Kibish 2; (c) Omo Kibish 1; (d) BOU VP-16/1/ Herto (after White et al. 2003 p 734)
of near-modern late archaics came from absolute dates for the Florisbad hominid from South Africa and for the Laetoli Hominid 18 from the Ngaloba Beds in Northern Tanzania (Manega 1995; Grün et al. 1996). The Florisbad specimen directly dated by ESR to ca. 250 Kyr BP has a coronally greatly expanded frontal bone associated with a continuous but only slightly projecting supraorbital torus and a modern facial shape with a well-developed canine fossa. The LH 18 cranium, with an age of more than 200 up to 300 Kyr, exhibits a modern-looking face with a canine fossa and a near-modern braincase with a capacity of about 1,350 cm³, a more or less rounded occipital bone, and well-developed parietal bossing. Archaic features mainly exist in the flat and narrow frontal squama and the supraorbital torus (Bräuer 1989). However, the relatively thick torus shows an incipient division in the midorbital region, which might be a tendency toward the fully modern pattern. Based on its combination of ancestral and derived conditions, the cranium from Eliye Springs, West Lake Turkana, might also belong to this grade (Bräuer and Leakey 1986; Bräuer et al. 2004a). Important late archaic crania also come from Jebel Irhoud, Morocco, dated to about 170 Kyr BP (Grün and Stringer 1991; Hublin 1992).

The mosaic-like transition from late archaic to early anatomically modern H. sapiens is also obvious in the available specimens from Ethiopia such as the Omo Kibish remains (Figure 14.3). New field work at the sites and new Ar/Ar dating has suggested that both the Omo 1 skeleton as well as the Omo 2 cranium date to 195 ± 5 Kyr BP (McDougall et al. 2005). Whereas the Omo 1 specimen is by most accounts fully anatomically modern and according to the new dating evidence the oldest known modern human, the Omo 2 cranium shows a mosaic of modern and archaic features. It exhibits a robust yet basically modern supraorbital morphology along with strong midsagittal keeling and an angulated occipital bone. Recently discovered cranial remains from Herto in the Middle Awash, Ethiopia, dated by Ar/Ar to ca. 154–160 Kyr BR, further illustrate this transitional process (White et al. 2003). The large robust Herto cranium BOU VP-16/1 (Figure 14.3) apparently exhibits a modern supraorbital morphology and modern face combined with a rather angulated occipital bone similar to the condition seen in Omo 2. The variability of this occipital trait, however, is evident in the less angled condition of the more fragmentary adult specimen BOU VP–16/2 (White et al. 2003). With an age of around 150 Kyr BR, the Singa cranium from Sudan also belongs to this earliest modern human spectrum (Bräuer 2001b).

More early moderns come from South Africa, especially from the Klasies River Mouth Caves. Here, the oldest human remains date to ca. 120 Kyr BP. These maxillary fragments fall within the range of variation of Holocene Africans (Bräuer et al. 1992b). A nearly complete mandible with an age of about
100 Kyr is anatomically modern, as are the other slightly more recent cranial fragments from the site (Bräuer 2001b). Regarding the postcranial specimens, the conditions seen in the Klasies remains can be matched with recent population samples from Southern Africa (Rightmire and Deacon 1991; Churchill et al. 1996). Some features, such as the relatively low coronoid height of the ulna, could be retained archaic features reflecting the mosaic pattern of evolution in the postcranial skeleton (Churchill et al. 1996). But it is also possible that such postcranial conditions simply belong to the range of variation of these early modern humans (Pearson 2000). Another early modern specimen from South Africa is the Border Cave 1 partial cranium, which is about 90 Kyr old or perhaps somewhat older (Grün and Beaumont 2001).

Considering the complete evidence from the fossil record, of which I can discuss only a few specimens here in some detail, it is obvious that there is good documentation of the modernization process in Africa. This was recently further supported by a new analysis of the Middle Pleistocene hominids showing clear trends from early archaic up to modern *H. sapiens* in many metrical and non-metrical cranial features (Mbuâ and Bräuer in press). The current fossil record of the African Middle Pleistocene allows subdivision into the three suggested groups, regardless whether they are termed “grades” of an evolving species *H. sapiens* or different “morphs” or “paleospecies.” Foley (2001) distinguishes these groups on the species level, as *H. heidelbergensis*, *H. helmei*, and *H. sapiens*. Yet he concedes that the derived descendant taxa of *H. heidelbergensis* are problematic because of the continuity that can be found between them and their presumed ancestor:

- These seem to be species in the sense that Simpson meant—lineages with independent trajectories—but both the details of the fossil record and the scale of the process seem to rule out any punctuated events. Indeed, continuity between them, rather than discontinuity, is the reason for the persistent problem of delimiting the taxonomic units in the later stages of human evolution and gives rise to the question of whether the species concept, which lies at the heart of macroevolutionary theory, is sufficiently fine-tuned to cope with evolution at this scale (Foley 2001 pp 9–10).

However, problems already exist with regard to *H. heidelbergensis*. Shortly after Adefris (1992) classified the Bodo cranium as “archaic *Homo sapiens*” based on a detailed study, Rightmire (1996) considered it most reasonable to refer to Bodo as a *H. heidelbergensis*. Rightmire (1998) suggested speciation between *H. erectus* and *H. heidelbergensis* in Africa at around 800–700 Ka, instead of a speciation between *H. erectus* and (archaic) *H. sapiens* (Bräuer 2001b). While this difference appears to mainly involve names, especially since Rightmire also favors
a single polytypic species *H. erectus* in Asia and Africa, his scheme also suggests two further speciations, one in Europe from *H. heidelbergensis* to *H. neanderthalensis* at ca. 300 Ka and another in Africa from *H. heidelbergensis* to *H. sapiens* at ca. 150 Ka. However, due to the mosaic nature of the accretion process in Europe, it is hardly possible to define any clear divisions along the Preneanderthal/Neanderthal lineage (Hublin 1998 p 302). Thus, there appears to be no convincing evidence for such a speciation within the European record. In Africa, there is a similar process with regard to the mosaic-like modernization, which also does not justify any subdivisions at the species level (Smith 2002; Turbón 2006). I think it is more likely that we are dealing with one evolving biological species and several subspecies (Jolly 2003). Further splitting of the African sequence as suggested by Foley and Lahr (1997) into three species (*H. heidelbergensis*, *H. helmei*, and *H. sapiens*), which largely include the same African specimens as the three proposed grades, can only mean that these are artificially defined morphs or paleospecies. Yet, it is problematic that Foley and Lahr (1997) also included European fossils in *H. helmei*. As outlined by Stringer (2002a p 567):

- First, Neanderthal characteristics were already evolving in Europe prior to the hypothesized appearance of “*H. helmei*”, e.g. in the Swanscombe specimen, dated to ca. 400 kyr. Second, African specimens such as Florisbad and Jebel Irhoud make unparsimonious ancestors for the Neanderthals, since not only do they post-date the appearance of Neanderthal clade characters in Europe, but they appear to lack Neanderthal morphological characteristics that might be expected in a common ancestor.

In spite of the current discussion of the number of taxa involved, which mainly arises from the different species concepts used, there is nearly general agreement that the evolutionary process toward modern humans in Africa can be divided into a sequence of three groups, leading to an early origin of modern humans during the late Middle Pleistocene. This is in agreement with the fact that in the Near East, anatomically modern humans appeared as early as about 100 or 130 Ka. These early moderns from Skhūl and Qafzeh might not only have an African origin but may also represent the oldest, well-documented evidence of modern humans outside of Africa.

### 14.3 Replacement of European Neanderthals

Largely parallel to the anatomical modernization process in Africa, the evolution of the Neanderthals took place in Europe. It is now widely agreed that the development of the Neanderthal morphology resulted from an accretion process,
which began about 500 Ka (Hublin 1998; Bischoff et al. 2003). Studies on Neanderthal mtDNA support such a long, separate evolution in Africa and Europe (Krings et al. 2000; Beerli and Edwards 2002). This does not necessarily indicate that two different biological species originated since the taxonomic significance of the genetic differences is ambiguous (Hublin 1998; Jolly 2003). So, whether one prefers to regard the Neanderthals as a subspecies of \textit{H. sapiens} (Turkón 2006) or as a different species or paleospecies (Tattersall and Schwartz 2000; Harvati 2003), the Neanderthals were clearly very closely related to us (Stringer 2001b). The long, separate evolution of the Neanderthals led to an accumulation of a large number of derived features in the skull and postcranial skeleton that are either unique or most frequent in this group (Trinkaus 1988; Hublin 1998).

In spite of the numerous differences between the Neanderthal morphology and that of early modern Europeans, multiregionalists have long regarded Central Europe as a possible region of evolutionary change from Neanderthals to modern humans (Smith 1982; Frayer et al. 1993). In particular, the Neanderthal remains from Vindija, Croatia, were considered to demonstrate evolutionary continuity to early modern humans of the region, as e.g., from Velika Pećina. However, direct AMS radiocarbon dating has shown that the frontal fragment from Velika Pećina once regarded to be ca 34 Kyr old is only 5 Kyr old and the most recent Vindija Neanderthal sample (G1) has an age of only 28,000 or 27,000 radiocarbon years BP (Smith et al. 1999). Based on a new ultrafiltration technique for bone radiocarbon samples an age of ca. 32 Kyr could also be possible for the Vindija G1 fossils (Higham et al. 2006). This technique will certainly be applied to other specimens as well in the future. Nevertheless, Neanderthals not only existed relatively recently on the Iberian Peninsula but also in Central Europe. AMS dating has led to further drastic revisions regarding some presumably early modern Europeans. The frontal bone from Hahnófersand near Hamburg, which was dated to ca. 36 Kyr (Berger and Protsch 1989 p 64), has been redated to only 7,500 years BP (Terberger et al. 2001). For the fragmentary crania from Zlatý Kůn and Svitávka, Czech Republic, for which ages of at least 30 Kyr were suggested, new dates of 12,870 and 1,180 years BP, respectively, were obtained (Svoboda et al. 2002). Also, the human remains from the Vogelherd Cave in southwestern Germany, which have long been thought to derive from Aurignacian deposits, turned out to be only 5 Kyr old (Conard et al. 2004).

In spite of the exclusion of several presumably early modern Europeans, there are still a number of specimens that date to more than 30,000 radiocarbon years BP. Among these are the remains from Mladeč, Czech Republic, for which direct AMS dating on several specimens provided ages of about 31 Kyr BP, which are in agreement with the Aurignacian artifacts as well as previous AMS dates for associated calcite deposits (Svoboda et al. 2002; Wild et al. 2005). Further early
modern cranial and mandibular remains have recently been discovered in the Peștera cu Oase, Romania, for which direct AMS dating of a mandible yielded ages of 34–36 Kyr BP (Trinkaus et al. 2003). Most recent AMS dating on animal bones associated with the presumably modern maxillary fragment Kent’s Cavern 4 from England has even suggested dates between 37 and 40 Kyr BP (Stringer 1990; Stringer personal communication). In spite of uncertainties of radiocarbon dates due to fluctuations in atmospheric radiocarbon (Conard and Bolus 2003), there is still good evidence for several thousand years of Neanderthal–modern coexistence. Thus, the question of a Neanderthal contribution to the early modern gene pool remains relevant and has been strongly debated between supporters of the Multiregional Evolution and Out-of-Africa models for many years.

Based on the current dates, the Mladecˇ remains represent the best preserved early modern material from the time at which Neanderthals were still around in Central Europe. Thus, it is key material for examining the question of Neanderthal–modern continuity or discontinuity. Multiregionalists have long used the Mladecˇ material to demonstrate regional evolution. Frayer (1986 p 254) saw here “good evidence in support of a gradualist model,” and Wolpoff (1999 pp 762–763) described the Mladecˇ specimens 4, 5, and 6 as Neanderthal-like. In testing the similarities of Mladecˇ 5 and 6 to their potential ancestors—European Neanderthals and early moderns from Skhûl/Qafzeh—Wolpoff et al. (2001 p 293) even found a predominance of Neanderthal resemblances for several sets of variables, which are said to separate the Neanderthal and Skhûl/Qafzeh samples.

In order to determine the hard evidence for the Neanderthal contribution to these early moderns, new analyses of the Mladecˇ material were carried out (Bräuer and Broeg 1998; Bräuer 2006; Bräuer et al. 2006). The new studies included the most relevant and well-preserved crania, the females Mladecˇ 1 and 2 and the males Mladecˇ 5 and 6 (Figure 14.4) as well as the maxilla fragment, Mladecˇ 8. A set of nonmetrical Neanderthal features was used as well as measurements describing the frontal shape and facial morphology and projection. The results were rather unexpected. Regarding the common nonmetrical Neanderthal features studied, not a single condition could be found in the Mladecˇ specimens that can be said to have unequivocally derived from Neanderthals. On the contrary, the respective regions, such as e.g., the supraorbitals, the shape of the vault, the face, the occipital, and mastoid regions, are all anatomically modern and do not exhibit any derived Neanderthal morphology. Also, the metrical analysis revealed great deviations in the frontal curvature of the Mladecˇ specimens from both European and Near Eastern Neanderthals. Among the Upper Paleolithic sample (n = 18), Mladecˇ 1 and 5 show the greatest divergence from Neanderthals (n = 10). No indication of a midfacial projection could be found in these early Czech specimens. Mladecˇ 1, 2, and 5 diverge strongly from
Neanderthals with regard to their Nasiofrontal Angle (NFA), and a specimen only for which the Zygomatic Angle (SSA) could be determined, shows a large disparity when compared to the Neanderthal range of variation.

Major reasons for these contradictory results lie in the assessment as well as the selection of the features. For example, one of the commonly used features is the “occipital bun” (chignon), which exhibits a derived morphology in Neanderthals. In early modern Europeans, the occipital projections are not only placed “in the context of a rather different cranial shape” compared to that of Neanderthals (Churchill and Smith 2000 p 97) but they also differ in a number of special details such as the extent of lambdoidal flattening, the angulation between the nuchal and occipital planes, and the height of the vertical face of the posterior occiput (Caspari 1991). These differences led Smith (1982) to suggest the term “hemibun” for the occipital projections in Upper Paleolithic specimens. Lieberman et al. (2000 p 291) even raised the possibility that the specialized Neanderthal bun and the projections in early modern Europeans may not be homologous. In addition, these more moderate projections are rather variable and occur practically all over

![Figure 14.4 Early modern crania from Czech Republic: (a) Mladeč 1; (b) Mladeč 2 (c) Mladeč 5; (d) Mladeč 6 (photos c, d: H. Broeg)]
the world as, e.g., in Africa or East Asia (Bräuer and Broeg 1998; Wu 1998 p 282). Thus, to assess as an “occipital bun” anything different from the derived and well-defined Neanderthal bun must obviously lead to confusion, as can be well demonstrated in the case of the Mladeč specimens.

Wolpoff (1999 p 762) classifies the occipital projections in Mladeč 5 and 6 as a “Neanderthal-like occipital bun,” a view shared by Frayer (1992 p 21). However, the midsagittal curvatures of Mladeč 5 and La Chapelle-aux-Saints are in fact more different than those of, e.g., Mladeč 5 and late Middle Pleistocene archaic H. sapiens specimens like Jebel Irhoud 1 or Laetoli H18 (Bräuer and Broeg 1998; Bräuer 2006). Smith et al. (1995 p 201) had also emphasized that “the occipital buns in Jebel Irhoud are more similar to those in early modern Europeans (like Mladeč, Zlatý Kůn, or Predmosti) than to Neanderthals because of the greater similarity in posterior cranial vault morphology between them.” Thus, there should be little reason to suggest that the more African-like hemibuns in Mladeč were derived from European Neanderthals. However, Frayer (1992 p 21) argued that the morphology in the early modern Europeans must have derived from Neanderthals because occipital buns are not present in the Near Eastern early moderns. Yet, in regard to the bunning in Jebel Irhoud and other circum-Mediterranean specimens, Wolpoff (1999 p 613) considers the bunning similar in Skhül/Qafzeh. And indeed, comparisons of the midsagittal curvatures demonstrate close similarities between the occipital profiles of Qafzeh 6 and those of all three Mladeč crania in which the occipital is preserved (Bräuer 2006). Thus, the predominant multiregionalist assessment of the bun morphology in Neanderthals and Mladeč as present and in Skhül/Qafzeh as absent is inaccurate.

Another example of problematic assessment concerns the “suprainiac fossa,” which also has a clearly defined morphology in Neanderthals. It is a generally large, wide resorptive depression which is either rectangular or triangular in shape with a horizontal base formed by the transverse occipital torus (Nara 1994; Hublin 1998). Yet, as Caspari (1991 p 184) pointed out, “it is unclear if all resorptive pitting in this area that could be referred to as fossae (or incipient fossae) can be validly compared with the Neandertal condition. Perhaps only the Neandertal pattern should be considered true suprainiac fossae, or alternatively some intermediate forms might also be considered suprainiac fossae.” Quite obviously, not all pitting in this region can be regarded as a Neanderthal suprainiac fossa. Smaller or larger resorptive depressions that do not match the Neanderthal condition can be found in modern humans nearly everywhere in the world. They are frequently present, e.g., in terminal Pleistocene material from North Africa and occur in recent crania from New Guinea (Bräuer and Broeg 1998). These depressions normally have an inverted triangular shape with the apex pointing to the nuchal plane. Mladeč 5 shows such a small pitting, which can
hardly be regarded as Neanderthal suprainiac fossa, and also Mladeč 6 exhibits a tiny non-Neanderthal-like depression. This had also been contended by Frayer (1986 p 251), when he concluded that “neither Mladeč 6 nor 5 have a distinct suprainiac fossa.” However, later Caspari (1991) felt that there was a suprainiac fossa in Mladeč 6, and Frayer (1992 p 22), although emphasizing the differences between the Neanderthal and Upper Paleolithic conditions, recognized the structure in Mladeč 6 as a possible exception. This example also highlights the problematic use of character states in order to support the idea of continuity.

Wolpoff et al.’s (2001) recent finding of a great number of Neanderthal resemblances in two of the Mladeč specimens is largely based on features which are problematic with regard to both their assessment and phylogenetic relevance. Many of the nonmetric features used are, in fact, metric traits that were arbitrarily divided into two alternative conditions without any recognizable justification such as, e.g., thick parietal at asterion (>9 mm), broad frontal (>125 mm), long frontal (glabella–bregma length >113 mm), or long occipital plane (>60 mm). With regard to other features used, it is unclear how they can be properly scored as present or absent without a clear scoring system. Among these are: mastoid–supramastoid crests well separated, frontonasal suture arched, glabellar depression, glenoid articular surface flattened, and paramastoid crest prominent. In addition, many of the features used can hardly be regarded as distinguishing between Neanderthals and modern humans. Thus, it is not surprising that Wolpoff et al. (2001) found the most divergent numbers of differences between Mladeč 5 and two closely related early modern specimens from Skhūl, i.e., only 8 differences to Skhūl 4, but 23 to Skhūl 5. Yet, there are further problems with this study, such as the morphological representativeness of the specimens employed, the lack of relevant features, and the inappropriate method of pairwise difference analysis (Collard and Franchino 2002; Bräuer et al. 2004b). Thus, critical reexamination of the claimed evidence for a considerable Neanderthal contribution in the Mladeč material cannot be supported.

Another early modern specimen claimed to show evidence for gene flow from Neanderthals is the mandible from Peștera cu Oase. However, according to Trinkaus et al. (2003 p 11235), “the only feature that suggests Neandertal affinities is the lingual bridging of the mandibular foramen, a feature that is currently unknown among humans preceding Oase 1 other than the late Middle and Late Pleistocene members of the Neandertal lineage.” Yet, how reliable is this so-called H−O foramen for demonstrating Neanderthal–modern gene flow or continuity? As Smith (1978 p 327) found, there are neither individuals under 18 years of age from any modern sample examined nor any juvenile Neanderthal exhibiting the H−O type. This makes it likely that, as Smith (1978) mentioned, factors developing during the individual’s lifetime could also be responsible for the occurrence of
the trait. Since the morphology at the mandibular foramen can change during a lifetime, the feature cannot be regarded as a good genetic trait (Lieberman 1995 p 175). Moreover, its occurrence does not clearly signal continuity. According to Frayer (1992 p 31), the H-O type appears in 10 out of 19 Neanderthals (52.6%), in 4 out of 22 early Upper Paleolithic specimens (18.2%), and in 2 out of 30 late Upper Paleolithic specimens (6.7%). The four early Upper Paleolithic specimens, however, include the Neanderthal Vindija 207 and Stetten (Vogelherd) 1, which no longer belong to this group. Thus, there is no relevant occurrence of this feature in the early Upper Paleolithic. Moreover, most important in this respect is a study of terminal Pleistocene North African samples. Groves and Thorne (1999 p 253) found the bilateral H-O foramen in 22.2% of the Nubian material from Tushka and Jebel Sahaba, further showing that this trait is not necessarily derived from Neanderthals. Finally, the H-O foramen already occurred outside of Europe in H. erectus such as in OH 22 and Zhoukoudian H1. Thus, though it cannot be excluded, it is far from clear that this single possible Neanderthal trait present on one ramus of the Oase mandible really indicates gene flow from Neanderthals. There is also no unequivocal evidence for a Neanderthal contribution in the adolescent cranium Oase 2 that might be contemporaneous with the mandible. This specimen exhibits a whole set of derived modern features in the vault and face (Trinkaus et al. 2006).

Finally, the juvenile Gravettian skeleton from Lagar Velho, Portugal, has been suggested to exhibit aspects indicating Neanderthal ancestry (Zilhão and Trinkaus 2002). Yet, even the major feature that has been suggested to show a Neanderthal-like condition—the relatively short lower legs—is not unequivocal. It cannot be excluded that the condition also reflects some adaptation among these early modern populations to cold conditions before the Last Glacial Maximum (Stringer 2002b) or individual variation within this modern population (Tattersall and Schwartz 1999). In addition, the occurrence of Harris lines points to periods of disease or malnutrition. Moreover, if one takes the estimated adult Crural Index for Lagar Velho of 80.7% and the 95% confidence intervals (77.5–83.9) as provided by Ruff et al. (2002 p 380), there appears to be considerable overlap with early Upper Paleolithic Europeans as well. No matter whether some Neanderthal genes were involved in the critical specimens discussed here or not, it is difficult to escape the conclusion that the hard evidence points at best to small traces of interbreeding between Neanderthals and modern humans. No evidence for significant interbreeding can be demonstrated convincingly among the early modern Europeans. Thus, a replacement process probably accompanied by a small amount of gene flow as suggested by the Out-of-Africa model appears to be in good agreement with the current evidence from Europe.
14.4 Continuity or discontinuity in the Far East

The assumption of regional evolution in China and Australasia is mainly based on morphological features or sets of features. Weidenreich (1943) published a list of 12 so-called Mongolian traits that he regarded as clear evidence for a close relationship between Chinese *H. erectus* and living north Chinese. Among these are: midsagittal crest and parasagittal depression, metopic suture, Inca bone, certain “Mongolian” features of the nasal bridge and cheek region of the maxilla and zygomatic bone, shovel-shaped upper lateral incisors, and horizontal course of the nasofrontal and frontomaxillary sutures. Coon (1962 p 458) was also convinced that many common features existed between Peking Man and living northeast Asians. Finally, the multiregionalists, as well, see strong support for their model in north Asia. Accepting most of the features suggested by Weidenreich and some other authors, multiregionalists presented a rather long list of assumed regional Chinese traits (Wolpoff et al. 1984; Pope 1992; Wu 1992). Although some features are said to have changed through time, Wolpoff et al. (1984 p 435) were convinced that “all of these characteristics have much lower frequencies and are distributed discontinuously in regions other than China.”

Weidenreich (1943) also saw an evolutionary line leading from *Pithecantropus* via *H. soloensis* to the Australian aborigines of today. Some years before, Keith (1936) had already regarded the Solo hominids, just discovered, as representing the evolutionary stage linking *Pithecantropus* to the Australian aborigine. Wolpoff et al. (1984 pp 443–444) mentioned about 20 continuity features of the cranium, including flatness of the frontal in sagittal plane, distinct prebregmatic eminence, marked prognathism, persistence of the zygomaxillary ridge, eversion of the lower border of the malar, rounding of the inferolateral border of the orbit, prelambdoidal depressions, and lambdoidal protuberance. Most of the features are from Weidenreich (1943), Larnach and Macintosh (1974), and Thorne and Wolpoff (1981). Based on these assumed regional features, multiregionalists are convinced that Australasia shows a clear anatomical sequence between the earliest Indonesian *H. erectus* and the recent and living Australians uninterrupted by African migrants at any time (Frayer et al. 1993 p 21).

Since many of the regional features for the Far East were suggested in the first half of the twentieth century, reexaminations have been carried out by several authors over the last two decades. With regard to the suggested East Asian traits, Groves (1989 p 279) arrived at the conclusion that there is “little evidence for special likeness of modern ‘Mongoloids’ to *Homo erectus pekinensis,*” and identified many of the features as primitive retentions also found outside the region. According to Habgood (1992 p 280) “none of the proposed ‘regional features’ can
be said to be documenting ‘regional continuity’ in east Asia as they are commonly found on modern crania from outside of this region . . . , and are consistently found on archaic *Homo sapiens* and/or *Homo erectus* crania throughout the Old World.” In her analysis of 11 proposed East Asian continuity features, Lahr (1994) found that almost all of them occurred *more frequently* in recent samples from other parts of the world. Facial flatness was found to be most pronounced in final Pleistocene populations from Northern Africa. A recent detailed study by Koesbardiati and Bräuer (in preparation) focused on the middle and upper facial regions especially regarded as showing a great number of East Asian continuity features (Pope 1992; Frayer et al. 1993). The results, however, revealed that no variable showed a distribution as would be predicted by the Multiregional Evolution model. Instead, most of the features occur more frequently or are more pronounced among population samples from Africa, Europe, and Australia than in the Chinese sample. Some of the features show the highest frequencies among the Inuit of Greenland who, however, represent a relatively young population showing adaptations to extreme conditions (Howells 1993; Lahr 1995).

Also, the fossil record from China hardly supports regional continuity from *H. erectus* to modern people of the area. Even the problematic continuity features are lacking in archaic *H. sapiens* specimens. For example, the 200 Kyr old cranium from Jinniushan does exhibit neither midfacial flatness, a marked angulation of the maxilla and zygomatic bones when viewed from beneath, nor a flattened nasal saddle, a nondepressed nasal root, or lack wisdom teeth. Based on further features, even Pope (1992) has raised doubts whether this hominid, along with the Maba specimen, fits into the picture of regional evolution in China and suggested strong genetic influence from the West into China in the terminal Middle or Late Pleistocene. Moreover, the 150 to 200 Kyr old cranium from Dali (Figure 14.5) with its heavy supraorbital torus and other archaic features hardly reveals clear similarities to modern Chinese. Instead of a regional evolution to modern humans, it is likely that archaic humans with torus-like supraorbitals, as indicated by the Laishui skeleton, Hebei Province, were still present as recently as 30 or 40 Kyr BP (Etler 1996; Lü personal communication).

Further relevant evidence contradicting the idea of regional evolution in China comes from the earliest modern humans of the area, especially the specimens from the Upper Cave at Zhoukoudian and from Liujiang in southern China, probably dating to ca. 25–30 Kyr BP (Hedges et al. 1995; Etler 1996). Due to recent Uranium-series dates, Shen et al. (2002) have argued that the Liujiang specimen could be older, but there remain basic uncertainties as to the stratigraphic position of the human remains and thus to their age. Several studies
of the specimens concluded that their cranial shape differs from that of recent Chinese (Howells 1995; Stringer 1999). A new multivariate analysis of early modern Chinese crania (Bräuer and Mìmisson 2004) demonstrated that the Upper Cave 101 cranium exhibits the closest affinities with Upper Paleolithic Europeans and also Australians and recent Europeans. For the Liujiang cranium (Figure 14.5), greatest affinities were found with recent North Africans and Europeans as well as with Upper Paleolithic Europeans, terminal Pleistocene North Africans and sub-Saharan Africans. Thus, the current evidence does not support the view of a long regional evolution in China but rather favors an immigration of modern humans into the region and an emergence of the morphological characteristics of recent Chinese only during the last 20 or 30 Kyr.

Regarding Australasia, the central assumptions of the multiregionalists have also proven to be no longer tenable. In spite of the huge gap in time between the only H. erectus specimen with a well-preserved face, Sangiran 17, and the earliest Australians, Wolpoff et al. (1984 p 443) saw, especially in the face, a number of
continuity features such as the marked prognathism, the maintenance of large posterior dentitions throughout the Middle and Late Pleistocene, an eversion of the lower border of the malar, the persistence of a zygomaxillary ridge, and rounding of the inferolateral border of the orbit. Yet, the assumption of interregional gene flow as emphasized by multiregionalists would make it very unlikely that such combinations of features could have been maintained over a period of more than 1 Myr and during the claimed transformation process from the massive facial morphology of *H. erectus* toward that of modern humans (Nei 1998). In addition, Aziz et al. (1996) demonstrated that the earlier reconstruction of the Sangiran 17 cranium done by Wolpoff (Thorne and Wolpoff 1981) is incorrect in several respects. In the revised reconstruction (Figure 14.6), the degree of facial prognathism is not marked, but only moderate, and there is no eversion of the lower border of the zygomatic bone, which as a whole bulges laterally. Moreover, the zygomaxillary ridge could not be detected by Aziz et al. (1996 p 20), nor by myself on the original. There is also no rounding of the inferolateral border of the orbit (Baba et al. 2000 p 61). Finally, Baba et al.

Figure 14.6
*Homo erectus* from Indonesia and an early modern Australian: (a) Revised reconstruction of Sangiran 17 by Aziz et al. 1996 (photo: F. Aziz/H. Baba); (b) Ngandong; (c) Lake Mungo 3
(2000 p 62) demonstrated that “it does not make sense to compare the degree of facial and dental reduction between Sangiran 17 and Australians, because masticatory adaptation pattern is completely different from each other.” Thus, relevant continuity features which have been said to be present on Sangiran 17, recently dated to ca. 1.3 million years ago (Antón 2003), are simply not there or are the result of incorrect reconstruction. Also, in her comprehensive analysis of 20 assumed Australasian cranial features, Lahr (1994) arrived at the conclusion that most of the traits did not show the pattern proposed by the Multiregional Evolution model, and quite a number of them occur with high frequency in a robust terminal Pleistocene sample from North Africa, thereby supporting a functional interpretation of these traits.

Further arguments against regional evolution in Australasia have been provided by new dating evidence. According to ESR and Uranium-series methods, the late *H. erectus* specimens from Ngandong/Solo (Figure 14.6) might only be 30–50 Kyr old (Swisher et al. 1996). They would thus be contemporaneous to the fully modern skeleton Lake Mungo 3, southern Australia (Figure 14.6), for which several dating approaches yielded an age of more than 40 Kyr and perhaps even up to 60 Kyr BP (Thorne et al. 1999; Bowler and Magee 2000). Consequently, a regional evolution from late *H. erectus* of Java to these rather gracile modern humans from Australia can be practically excluded. In addition, the assumed greater age of a very robust cranium, WLH-50 from the Willandra Lakes (Wolpoff 1999 p 736), could not be supported by a recent absolute date of only 14 Kyr BP (Simpson and Grün 1998). Moreover, the cranial vault of this specimen has been modified by severe pathological alterations (Webb 1990). Notwithstanding the young age and pathology of WLH-50, Wolpoff et al. (2001) chose this cranium instead of the much more appropriate one from Lake Mungo in order to demonstrate regional evolution in Australasia. However, their approach also suffers from fundamental problems with the characteristics used as well as their assessments (Bräuer et al. 2004b).

Although no conclusive indications for a regional evolution from *H. erectus* to modern Australians appear to exist, the human fossil and subfossil material from Australia as a whole exhibits considerable variability with regard to robusticity. Among these anatomically modern humans, there are very robust as well as gracile specimens as, for example, present in the Kow Swamp material which dates to ca. 22–19 Kyr BP or somewhat more recent (Bräuer 1989; Stone and Cupper 2003). Such robust specimens also date from the early Holocene. Possible explanations for the great range of variation might be drift effects in the course of the occupation of the continent but may also include adaptation to an increasingly arid environment toward the peak of the last glacial period.
(Klein 1999; Baba et al. 2000). Whether gene flow between the Ngandong populations and expanding modern groups also contributed to the variability in the Late Pleistocene Australians must remain open in light of the current evidence.

14.5 Conclusions

Over the last two decades, the accumulating evidence has strongly supported the Out-of-Africa model, allowing for the possibility of gene flow during the replacement period. Claims by multiregionalists for a significant contribution of Neanderthals to the early modern Europeans cannot be supported by the current fossil evidence. Instead, the most likely scenario appears to be a replacement of the Neanderthals accompanied by only a small amount of gene flow. For the Far East, it is more difficult to estimate the extent of interbreeding since the fossil record of the replacement period is rather poor. Nevertheless, in view of the serious problems regarding the claimed continuity features for China and Australasia and the fact that the earliest modern humans in both areas do not fit into the concept of regional evolution, no convincing evidence can be seen for long-term regional evolution up to modern humans or for a significant archaic–modern gene flow.

The anatomical modernization process in Africa now appears to be well established by quite a number of diagnostic and reliably dated hominid specimens. To describe the obvious morphological changes of this continuously evolving lineage, three grades of *H. sapiens* have been suggested. This evolutionary sequence led to an early emergence of anatomically modern humans nearly 200 Ka. In view of the obvious continuity from archaics to early moderns and the near-modern morphology of the premoderns, it appears unlikely that anatomically modern humans differ from their direct ancestors on the species level (Stringer 2002a). Instead, during the major part of the Middle Pleistocene, the African fossil record documents a diachronic increase of more derived modern-like conditions. Therefore, it appears adequate to assign the post-*H. erectus* humans to one biological species *H. sapiens* (Klein 1999). This view is also supported by the fact that the Preneanderthal/Neanderthal lineage which goes back to around 500 Ka BP cannot reasonably be split into different species (Hublin 1998). It might thus be in agreement with the current evidence (Figure 14.7) from Africa and Europe to only assume a speciation between *H. erectus* and (archaic) *H. sapiens* at around 700 or 800 Ka in Africa. Later, early archaic sapiens populations spread into Europe evolving toward the Neanderthals. Regarding China, it remains unclear whether the late Middle Pleistocene humans derived
from an eastward migration of archaic *H. sapiens* populations or were the result of regional evolution in China, or even originated by both processes.

As discussed in this and other chapters, various alternative phylogenetic scenarios have been suggested to describe Middle and Late Pleistocene human evolution. *H. sapiens* is often restricted to modern humans, and its ancestral group (largely equivalent to late archaic *H. sapiens*) has been assigned either to the same species (Stringer 2002a) or to a separate species, *H. helmei* (McBrearty and Brooks 2000). Alternatively, *H. helmei* is regarded as not only ancestral to
modern *H. sapiens* but also to *H. neanderthalensis* (Foley and Lahr 1997). These various possibilities, however, have their own problems. For example, the latter phylogeny does not consider that Neanderthal features had long existed in Europe prior to the hypothesized appearance of *H. helmei*. Moreover, suggesting several species within the continuously evolving lineages of later human evolution causes problems in delimiting the taxonomic units (Foley 2001). Such units are rather arbitrarily defined entities without any clear relation to biological species. Studies on extant primates led Jolly (2003 p 662) to conclude that Neanderthals, Afro-Arabian “premodern” populations, and modern humans are, roughly speaking, biological subspecies, comparable to interfertile allopatric taxa or phylogenetic species of baboons. Based on research in contemporary zones of hybridization, as the Awash anubis-hamadryas hybrid zone, Jolly (2001 p 198) suggested “that unless an undocumented, radical genetic event occurred in the 600 ka since they shared mtDNA ancestry with the Neandertals, premodern humans were certainly able to interbreed with them and produce viable, fertile, offspring as hamadryas and anubis baboons do.” Thus, it appears well supported from different lines of evidence to regard the European Preneanderthals/Neanderthals and the African Middle Pleistocene lineage to modern humans and possibly the late archaic group in China as belonging to one polytypic species *H. sapiens*. “Speciation remains the special case, the less frequent and more elusive phenomenon, often arising by default” (Grubb 1999 p 164).

**Acknowledgments**

I thank Winfred Henke and Ian Tattersall for inviting me to contribute to this handbook. I am grateful to Fachroel Aziz, Hisao Baba, Helmut Broeg, and Tim White for kindly providing photographs. I also thank Anja Bodendiek, Juliette Kober, Hermann Müller, and Eszter Schoell for their helpful support with the final version of the manuscript. Many thanks go to Mark Collard for valuable indications.

**References**


Bräuer G (1984b) The “Afro-European sapiens hypothesis” and hominin evolution in East Asia during the late Middle and Upper Pleistocene. Cour Forsch Inst Senckenberg 69: 145–165

Origin of modern humans


Manega PC (1995) New geochronological results from the Ndutu, Naisiusiu and Ngaloba Beds at Olduvai and Laetoli in Northern Tanzania: their significance for evolution of modern humans. Bellagio Conference, Italy


Nara MT (1994) Etude de la variabilité de certains caractères métriques et morphologiques des Néandertaliens. Thèse de Docteur; Bordeaux


15 Analyzing Hominid Phylogeny

David Strait · Frederick E. Grine · John G. Fleagle

Abstract

An understanding of the phylogenetic relationships among organisms is critical for evaluating the evolutionary history of their adaptations and biogeography as well as forming the basis for systematics. As the numbers of hominid fossils and hominid taxa have increased over the past 40 years, controversies over phylogeny have expanded and have become a hallmark of paleoanthropology. Concordant with the rise in taxonomic diversity, the increased use of phylogenetic systematics, or cladistics, has provided a valuable tool for reconstructing hominid phylogeny. Despite the widespread view that hominid phylogeny is a source of endless debate, there is a broad consensus regarding many aspects of hominid phylogeny.

15.1 Introduction

Phylogeny is central to our understanding of virtually any aspect of an organism’s biology. An appreciation of the phylogenetic relationships of an organism not only provides a perspective on its place in the history of life and a basis for taxonomy but is also critical for evaluating biogeography as well as ecology and behavior. The adaptations of all organisms, even those as adaptively flexible as primates, are inherited, and thus any proper statistical analysis of physiological adaptations requires a consideration of phylogeny (Harvey and Pagel 1991; Purvis and Webster 1999).

In paleoanthropology at the beginning of the twenty-first century, the study of phylogeny hardly needs to be justified. The first questions that are asked about any new fossil discovery are “How is it related to us?” and “What does it tell us about our evolutionary past?” Paleoanthropology has a reputation in the popular press as a discipline characterized by disagreements over phylogeny, and this is perhaps not unfair if one considers the seemingly diverse array of phylogenetic hypotheses that have been proposed in the last two decades (Delson 1986; Walker et al. 1986; Chamberlain and Wood 1987; Grine 1988; Kimbel et al. 1988, 2004; Wood 1988, 1991, 1992; Skelton and McHenry 1992, 1998; White et al. 1994;
Leakey et al. 1995, 2001; Brunet et al. 1996, 2002; Lieberman et al. 1996; Strait et al. 1997; Strait and Grine 1998, 1999, 2001, 2004; Asfaw et al. 1999; Senut et al. 2001). Certainly much of the current interest in hominid phylogeny has been fueled by new paleontological discoveries. Eight new early hominid species have been described since 1994 (White et al. 1994; Leakey et al. 1995, 2001; Brunet et al. 1996, 2002; Asfaw et al. 1999; Senut et al. 2001; Haile-Selassie et al. 2004), and our knowledge of the morphology of many previously described taxa continues to increase. Many of these discoveries have been accompanied by phylogenetic hypotheses, not all of which are compatible with each other. However, the current importance of phylogeny in paleoanthropology and our current understanding of hominid phylogeny are not just the result of new fossil discoveries. They also reflect major theoretical and methodological advances in our discipline during recent decades (Tattersall 1999). Thus, before discussing current views on hominid phylogeny, we provide a brief history of this endeavor during the last century. We use the term hominid in the traditional sense to mean taxa that are more closely related to humans than to any other primate.

15.2 Phylogenetic diversity in hominid evolution

As numerous authors have emphasized, theoretical approaches to hominid phylogeny changed considerably through the course of the twentieth century (Fleagle and Jungers 1982; Tattersall 1999; Gundling 2005). In the early decades, discussions of hominid phylogeny were largely limited to evaluating whether the few fossil taxa that were known at the time—*Pithecanthropus erectus* from Java, *Cyphanthropus rhodesiensis* or Rhodesian Man from Africa, Piltdown (*Eoanthropus dawsoni*) from England, Neandertals from Europe and, after 1925, *Australopithecus* and allied forms from Africa—were ancestral to living humans in their various forms. Two distinct issues dominated the literature. First, were any of the various fossil forms directly in the lineage leading to modern humans? As noted by Dobzhansky in 1944, most authorities found that all fossils had features which precluded placing them directly in human ancestry so that phylogenetic trees generally show a main trunk leading from somewhere in our primate past to modern humans, with each fossil taxon occupying a side branch leading to extinction (Figure 15.1; Dobzhansky 1944; Tattersall 1999). Despite the lack of reliable estimates of the geological age of any extinct taxa, they were generally suggested to have branched from the (main) human lineage at different times, such that each documented some aspect of human ancestry. However, all discussion concerned the relationship of the extinct taxa to the main human lineage, with little discussion of the relationships among the extinct taxa themselves.
Figure 15.1
A phylogeny of Primates from Hooton (1931) showing a tree with human races at the crown and the few known fossil taxa as long side branches from the main stem.
The notable exception to this was Weidenreich’s trellis model of hominid evolution in which all living and extinct taxa were interconnected but with temporal and geographic differentiation (Weidenreich 1946; Smith 1997). To a large degree, these trees showing humans at the crown just reflect the fact that in the early part of the twentieth century, as today, paleoanthropology was different from other aspects of zoology in being by definition focused primarily on tracing the history of a single organism, ourselves, rather than on the interrelationships of a large group of more or less equally important taxa. After all, it is our species that writes the books.

However, as Gundling (2005) has argued, a related but distinct, and in many cases more important issue in the first half of the twentieth century was the taxonomic issue of whether the various extinct species were hominids or apes. That is, where should the ape–human boundary be drawn? In most cases, these two approaches yielded concordant views. Fossils that were placed on the ape lineage were clearly not hominids. However, in some cases fossil taxa might be considered offshoots of the main human stem but still considered apes because they lacked the critical hominid character. This was at the heart of much of the debate regarding the place of Piltdown and *Australopithecus* in hominid phylogeny during the early part of the twentieth century. In general, most researchers limited the human family to modern people.

As Tattersall (1999) has so eloquently discussed, all of this changed in mid-century as paleoanthropologists began, however slowly, to adopt the tenets of the New Darwinian Synthesis. Despite a growing record of new fossils, human evolution was increasingly seen as a unilinear progression through time, with all morphological diversity consigned to intraspecific variation due to geography or sexual dimorphism (Buettner-Janusch 1966; Brace 1967). All of the divergent branches from earlier in the century were incorporated into the main stem, and human evolution was seen as one continuous chain of forms separated mainly by time. The most extreme expression of this approach was Mayr’s (1950) inclusion of all fossil hominids (including “robust” and “ gracile australopithecines”) into a single genus, *Homo*, with three species. Certainly there were other views, for example Robinson, following Broom, repeatedly argued that *Paranthropus* was a separate lineage of hominid from *Australopithecus* and *Homo*, and Louis Leakey argued from time to time that the ancestry of the human lineage was not to be found among known fossils of the time. However for much of the discipline, there was little appreciation of phyletic diversity in human evolution (Figure 15.2). This view of limited phyletic diversity was very compelling. It was supported by the leading authorities on evolutionary biology at the time, such as Mayr and Dobzhansky; it brought paleoanthropology in line with the rest of evolutionary biology; and it conformed well with what we see in the world today:
a single species of humans with considerable intraspecific variation. Moreover, there were theoretical reasons offered to justify a lack of phyletic diversity in a culture-bearing creature (Mayr 1950; Wolpoff 1971). And while extrapolating human behavior backward into the fossil record may not be totally justified, especially for the Pliocene, it is not totally unreasonable.

![Figure 15.2](chart.png)

A chart of hominid evolution from a major textbook from 1966 (Buettner-Janusch 1966) showing the conservative lumpers’ view on the left and the extreme splitters’ view on the right.

<table>
<thead>
<tr>
<th>Epoch</th>
<th>Fossils</th>
<th>Number of skulls</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Alternative classifications</td>
<td></td>
</tr>
<tr>
<td>Pleistocene</td>
<td>Homo sapiens</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Homo erectus</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Australopithecus</td>
<td></td>
</tr>
<tr>
<td>Miocene</td>
<td>Ramapithecus punjabicus</td>
<td></td>
</tr>
<tr>
<td>Upper</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lower</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Middle</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Upper</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Upper</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

10-20 Skulls
However, by the 1970s the unilinear view of human evolution was being seriously challenged on several fronts and had become increasingly difficult or even impossible to support. There were clearly two distinct hominid lineages present in the Late Pliocene and Early Pleistocene at Olduvai Gorge and Koobi Fora (Leakey and Walker 1976), and there was increasing evidence that in other parts of the world modern humans had preceded or were contemporary with European Neandertals (Leakey 1969; Stringer 1974, 1978; Howells 1975; Bräuer 1982). Likewise, “new” analytical approaches emphasized the view that morphological changes over the past 3 Myr or so did not follow a simple temporal pattern of increasingly modern features through time (Eldredge and Tattersall 1975).

With an increasing number of contemporaneous taxa, the potential phylogenetic complexity of the hominid fossil record continued to grow, and came to a head with the description of *Australopithecus afarensis* (hereafter called *Praeanthropus afarensis*) in 1978, and the ensuing debate over taxonomic diversity and phylogenetic relationships in early hominid evolution (Johanson et al. 1978; Johanson and White 1979; Tobias 1980; Olson 1981, 1985; White et al. 1981; Rak 1983; Kimbel et al. 1984; Skelton et al. 1986). The level of debate over early hominid diversity and phylogeny was heightened even further with the discovery in Kenya of the Black Skull (KNM-WT 17000) several years later (Delson 1986; Walker et al. 1986; Grine 1988; Kimbel et al. 1988; Wood 1988). Similarly, debate over the timing and geography of modern human origins and the relationship between *Homo sapiens*, Neandertals, and *Homo erectus* expanded in the 1980s and has yet to abate (Smith and Spencer 1982; Mellars and Stringer 1989; Trinkaus 1989; Stringer 2002).

### 15.3 Reconstructing phylogeny: the rise of cladistics

The increasing evidence of taxonomic and phyletic diversity in hominid evolution during the 1970s and 1980s coincided with the increasing prominence of phylogenetic systematics or cladistics in paleoanthropology (Eldredge and Tattersall 1975; Luckett and Szalay 1975; Delson et al. 1977; Tattersall and Eldredge 1977; Delson 1985; Skelton et al. 1986; Wood et al. 1986; Grine et al. 1987; Grine 1988). The methods of phylogenetic systematics, or cladistics, were developed by the German entomologist Willi Hennig in 1950, but it was only with their publication in an English translation in 1966 that his methods became widely known and applied in morphological studies to understanding the phylogeny of all sorts of organisms.

Cladistics is a method of phylogenetic reconstruction premised on the notion that not all morphological similarities are indicative of phylogeny (see...
also Volume 1 Chapter 5). Rather, only those similarities that are derived (i.e., novel) and inherited from a recent common ancestor should be indicative of patterns of relatedness. In practice, it is difficult (if not impossible) to discern, a priori, such features (called synapomorphies) from other types of similarities such as primitive retentions (symplesiomorphies), or traits that have evolved convergently or in parallel (homoplasies). Thus, cladistics relies on the principle of parsimony to identify synapomorphies and, hence, to reconstruct phylogeny. In a general sense, parsimony is the idea that the simplest explanation is the best one because it makes the fewest assumptions. As applied to cladistics, parsimony dictates that the best cladogram is the one that requires the fewest number of homoplasies or independent appearances of the same feature (see also Volume 1 Chapter 6). Parsimony analysis is conventional in evolutionary biology (Kitching et al. 1998) but is viewed with skepticism by some paleoanthropologists (Trinkaus 1990; Asfaw et al. 1999; Hawks 2005). This skepticism is misplaced because, at its core, the logic of parsimony is intuitive and not too dissimilar from that of “traditional” evolutionary systematics (e.g. Olson 1981). More significantly, it provides a replicable criterion for evaluating alternative hypotheses beyond preconceived notions of how things should be (Tattersall 1996).

Consider an example in which a phylogenetic analysis is being performed on the living hominoids (Hylobates, Pongo, Gorilla, Pan, and Homo) and a fossil hominid (Australopithecus). Now consider a character with two states (knee joint valgus or varus). The nonhuman apes and various outgroup taxa (other Old World higher primates) have a varus knee, while Homo and Australopithecus have a valgus knee. Given a cladogram in which Australopithecus and Homo are sister taxa (> Figure 15.3a), what can be concluded about the evolution of the knee joint? There are actually many ways in which the knee joint might have evolved. It is possible that a valgus knee joint was present in all of the ancestors represented by the internal nodes of the cladogram (> Figure 15.3b). Such a reconstruction requires that a varus knee joint evolved in parallel in each nonhuman ape lineage. Alternatively, it is possible that a varus knee joint was present at all of the nodes of the cladogram, including the one representing the last common ancestor of hominids (> Figure 15.3c). This reconstruction requires that a valgus knee joint would have evolved in parallel in Homo and Australopithecus. Neither of these reconstructions is satisfying because both are needlessly complex. There is instead a much simpler (i.e., more parsimonious) explanation for the evolution of the knee joint, namely that a valgus knee joint evolved once in the last common ancestor of the hominids Australopithecus and Homo, who subsequently passed that trait onto its descendants (> Figure 15.3d). Such a reconstruction does not require any homoplasy.
Figure 15.3
Principles of cladistics. (a) Cladogram depicting possible phylogenetic relationships among hominoids. (b) Pattern of character evolution in which a varus knee evolves many times in parallel in each of the nonhuman hominoids. (c) Valgus knee evolves in parallel in *Homo* and *Australopithecus*. (d) Most parsimonious pattern of character evolution in which a valgus knee evolves once in the last common ancestor of *Australopithecus* and *Homo*. (e) Cladogram depicting an alternative phylogeny. (f) Most parsimonious pattern of character evolution in the alternative cladogram. (a–d) is preferred over (e) because it involves fewer changes; and it is therefore the most parsimonious.
Now consider an alternative cladogram in which *Australopithecus* is the sister taxon of all of the living hominoids (Figure 15.3e). In this tree, the most parsimonious reconstruction of the evolution of the knee joint is one in which a valgus knee evolved in parallel in *Australopithecus* and *Homo* (Figure 15.3f). No other possible reconstruction of character evolution in the knee joint requires fewer character state changes or steps. Now consider that the two cladograms presented here (Figure 15.3a and e) represent alternative interpretations of hominoid phylogeny. How can these cladograms be compared so as to select one of them as the better hypothesis of phylogeny? Parsimony states that the preferred cladogram is the one that is simplest, namely, the one that minimizes the number of homoplasies required. Fewer homoplasies are required in Figure 15.3d than in Figure 15.3f, so the preferred cladogram is the one in which *Australopithecus* and *H. sapiens* are sister taxa (Figure 15.3a). There is nothing controversial about this example, and both cladists and noncladists would agree with the result. The only difference between this example and an actual cladistic analysis is that most analyses would examine many characters at once. This is a great advantage of numerical cladistic analysis over “traditional” evolutionary systematics in which only a handful of characters tends to strongly influence the shape of phylogenetic trees. Even more significant is the fact that cladistic studies make explicit assumptions and predictions so that analyses are replicable, and the results are testable.

### 15.4 Cladistic analyses of hominid phylogeny

#### 15.4.1 Early studies

The first cladistic analysis of hominid evolution was by Eldredge and Tattersall (1975) who also coauthored a series of papers delineating various levels of phylogeny reconstruction from producing a cladogram, to creating a phylogenetic tree, and finally an evolutionary scenario (Delson et al. 1977; Tattersall and Eldredge 1977). In the late 1970s and early 1980s, cladistic analyses in paleoanthropology were relatively simple and often consisted of little more than producing a cladogram and identifying a few shared derived characters at each node (Olson 1978; Andrews 1984; papers in Luckett and Szalay 1975; Delson 1985; Wood et al. 1986; Grine et al. 1987). Nevertheless, this was major advance from much previous work in primate phylogeny in that there was a clear effort to distinguish shared derived features from shared primitive ones, and authors provided explicit morphological justification for phylogenetic grouping at every level. Falsifying a set of relationships based on a cladistic analysis generally
requires identification of additional morphological features that produce a different cladogram when analyzed. As Tattersall (1999) has pointed out, the rise of cladistics has led to a tremendous increase in the detailed documentation and analysis of hominid morphology.

The mid-1980s saw the first use of quantitative cladistic analyses in hominid evolution. By using computer algorithms researchers were able to evaluate dozens of characters and compare hundreds of trees, tasks that were simply unfeasible otherwise. The first efforts to evaluate hominid phylogeny using numerical methods were in an analysis of early hominid phylogeny by Chamberlain and Wood in 1987 and a study of the genus Homo by Stringer (1987). Numerous subsequent analyses of the phylogeny of hominids and many other groups of primates have used essentially the same methods (Fleagle and Kay 1987; Kay et al. 1997; Strait et al. 1997; Ross et al. 1998).

15.4.2 Early hominid phylogeny

The 1987 study of early hominid phylogeny by Chamberlain and Wood did not include *Paranthropus aethiopicus*, which subsequently became the linchpin of early hominid phylogeny. The first study to include this species was that of Wood (1988), who examined the trait list provided by Walker et al. (1986) in their description of KNM-WT 17000. Wood (1988) found that *Paranthropus robustus* and *Paranthropus boisei* are sister taxa, that *Homo* is the sister taxon of this clade, and that *Praeanthropus afarensis, Paranthropus aethiopicus* and *Australopithecus africanus* branch off in sequence from the base of the hominid tree (Figure 15.4a). Notably, the three “robust” species are paraphyletic. A subsequent study by Skelton and McHenry (1992), using a more extensive trait list, found an identical cladogram. Wood (1991, 1992), using a data set composed entirely of craniometric measurements, found a most parsimonious cladogram in which *Australopithecus africanus* is the sister taxon of *Homo*; *H. habilis sensu stricto* and *H. rudolfensis* are sister taxa; and *Paranthropus boisei* and *Paranthropus robustus* are monophyletic (Figure 15.4b). Technically, Wood’s (1991, 1992) cladogram does not include *Paranthropus aethiopicus*, but it is reported in the text of his analysis that this species is the sister taxon of *Paranthropus boisei*. Lieberman et al. (1996) found a most parsimonious tree (Figure 15.4c) in which *Paranthropus* is paraphyletic, *Paranthropus robustus* and *Paranthropus boisei* are sister taxa, and *Australopithecus africanus* is nested within the *Homo* clade. Subsequently, Strait et al. (1997) found a cladogram in which *Paranthropus* is monophyletic and the sister taxon of *Homo* (Figure 15.4d). Recently, an independent analysis by Kimbel et al. (2004) has largely corroborated Strait et al.’s
(1997) results. Kimbel et al. (2004) found two equally parsimonious trees; one is equivalent to those of Strait et al. (1997), and the other differs only in placing *Australopithecus africanus* as the sister taxon of the *Parathropus* clade.

The analyses noted above appear to differ from each other, but in fact they are similar to a much greater degree than is generally acknowledged. The results of Strait et al. (1997) and Kimbel et al. (2004) differ from those of Wood (1988) and Skelton and McHenry (1992) only with respect to the relationships of *Paranthropus aethiopicus*. They differ from those of Wood (1991, 1992) principally with respect to the relationships of *Australopithecus africanus*. The
most parsimonious tree of Lieberman et al. (1996) differs from that of Wood (1988) and Skelton and McHenry (1992) only with respect to *Australopithecus africanus*. Thus, these cladograms disagree primarily with respect to only two taxa, *Australopithecus africanus* and *Paranthropus aethiopicus*. There are also disagreements concerning the exact relationships of *H. habilis* and *H. rudolfensis*, but all analyses that include these taxa place them at the base of the *Homo* clade. In short, it appears as if cladistic analyses of early hominids are converging on a common set of relationships. It would be an overstatement to claim that the pattern of early hominid phylogeny is known, but insofar as repeatability is a key component of any scientific result, it would appear that the broad strokes of early hominid phylogeny are perhaps better understood than commonly acknowledged.

### 15.4.3 New early hominid discoveries

There have been many discoveries of new fossil hominid species in the past decade (White et al. 1994; Leakey et al. 1995, 2001; Brunet et al. 1996, 2002; Asfaw et al. 1999; Senut et al. 2001; Ward et al. 2001). Despite the common refrain that phylogenetic debates can be resolved by the discovery of new fossils, many of these new finds raise rather than resolve phylogenetic questions. Although most new discoveries are accompanied by a phylogenetic hypothesis, those hypotheses often only address the relationships among only a few hominid taxa. Moreover, these hypotheses are potentially difficult to test using cladistic analysis because they specify ancestor–descendant relationships without specifying sister-group relationships. The remains of many of these new hominid taxa have not yet been thoroughly published, and further documentation of their morphology will doubtless permit more thorough analyses. However, on the basis of information available so far, we can use cladistic analysis to test many of the initial hypotheses that have been proposed because phyletic relationships imply sister-group relationships. In particular, it is an accepted principle that a species can only be an ancestor of another taxon if it is the sister species of that taxon, and if its character states resemble those reconstructed as being present in the relevant internal node of a cladogram (Szalay 1977; Smith 1994; Wagner and Erwin 1995; O’Keefe and Sander 1999). Accordingly, the phyletic hypotheses that have been proposed for many recent fossil discoveries in hominid evolution can be evaluated through reconstruction of sister-group relationships (Figure 15.5).

Strait and Grine (2004) have recently tested many of the hypotheses generated by these new taxa. They found that the pattern of hominid phylogeny is
unbalanced such that many species branch off by themselves from the base of the tree, while the top of the tree is dominated by two multispecies clades, *Homo* and *Paranthropus* (Figure 15.6). *Sahelanthropus* and *Ardipithecus* are, respectively,
the first two branches of the tree, with subsequent branches successively represented by *Australopithecus anamensis*, *Praeanthropus afarensis*, *Australopithecus garhi*, and *Australopithecus africanus*. If *Kenyanthropus platyops* is a valid species then its position within the *Homo + Paranthropus* clade is unresolved; it is either the sister taxon of the rest of the clade or of *Paranthropus*. Relationships within *Paranthropus* are also unresolved, with *Paranthropus boisei* being the sister taxon of either *Paranthropus aethiopicus* or *Paranthropus robustus*. The position of *H. habilis* relative to *H. rudolfensis* is unresolved, but it is clear that one or the other is the basal branch of the *Homo* clade. *Homo ergaster* and *H. sapiens* are sister taxa. These results are consistent with certain of the hypotheses offered in the original descriptions and inconsistent with others.

### 15.4.3.1 *Ardipithecus ramidus*

At the time of its description (White et al. 1994), *Ardipithecus ramidus* was the oldest and most morphologically primitive hominid species then known. White et al. (1994) suggested that *Ardipithecus ramidus* lies near the ancestry of all other hominids and that it may be the actual ancestor of those species. A cladogram consistent with this hypothesis would place *Ardipithecus ramidus* as the sister taxon of a clade that includes all other hominid species (Figure 15.5a). Strait and Grine (2004) found that *Ardipithecus ramidus* is the sister taxon of all hominids except *Sahelanthropus* (Figure 15.6). These results support the
hypothesis of White et al. in a general sense insofar as *Ardipithecus ramidus* branches off near the base of the hominid tree, if not necessarily at the basal node.

**15.4.3.2 Australopithecus anamensis**

The following year, Leakey et al. (1995) described *Australopithecus anamensis* as a species intermediate both chronologically and morphologically between *Ardipithecus ramidus* and *Praeanthropus afarensis*. Leakey et al. (1995; Ward et al. 2001) have suggested that *Australopithecus anamensis* is more closely related to later hominids than is *Ardipithecus ramidus* and may be directly ancestral to *Praeanthropus afarensis* (see also Volume 3 Chapter 7). A cladogram consistent with this hypothesis would depict *Australopithecus anamensis* as diverging from a higher node on the hominid tree than *Ardipithecus ramidus* and as the sister taxon of all later hominids (**Figure 15.5b**). An alternative topology that might also be consistent with the phyletic hypothesis would make *Australopithecus anamensis* the sister taxon of *Praeanthropus afarensis*. Strait and Grine’s (2004) results (**Figure 15.6**) are consistent with the hypothesis that *Australopithecus anamensis* is the sister taxon of all hominids except *Ardipithecus* (and, presumably, *Sahelanthropus*).

**15.4.3.3 Australopithecus bahrelghazali**

The discovery of *Australopithecus bahrelghazali* was notable primarily because it represented the first early hominid species found in central Africa. Brunet et al. (1996) did not offer a detailed phylogenetic hypothesis for *Australopithecus bahrelghazali* but rather noted merely that the species is more derived than the contemporaneous *Praeanthropus afarensis*. Not all workers accept that *Australopithecus bahrelghazali* and *Praeanthropus afarensis* are distinct species (Kimbel et al. 2004). This species was not included in the analysis by Strait and Grine because it is known from only a few remains.

**15.4.3.4 Australopithecus garhi**

As described by Asfaw et al. (1999), *Australopithecus garhi* preserves an unexpected combination of cranial and dental characteristics. In particular, it has megadont molars and premolars but a relatively primitive-appearing face and neurocranium. Asfaw et al. (1999) implied that *Australopithecus garhi* could
be a suitable ancestor for *Homo*, although they noted that the exact phylogenetic relationships of this species remained unresolved. They presented a cladogram in which *Australopithecus garhi*, *Australopithecus africanus*, *Paranthropus robustus*, *Paranthropus boisei*, *Paranthropus aethiopicus* and *Homo* form a clade but in which relationships within that clade were left unresolved. However, they presented four phyletic trees, and in three of those, *Australopithecus garhi* was posited to be an ancestor of at least some members of the genus *Homo*. Moreover, they (Asfaw et al. 1999 p 632) state that “If *A. garhi* proves to be the exclusive ancestor of the *Homo* clade, a cladistic classification would assign it to genus *Homo*.” Such a classification would only be valid if *Australopithecus garhi* and at least some of the *Homo* species form a monophyletic group, as in Figure 15.5c. Furthermore, in reference to the morphology of *Australopithecus garhi*, Asfaw et al. (1999 p 634) state that “its lack of derived robust characters leaves it as a sister taxon to *Homo* but absent many derived *Homo* characters.”

Strait and Grine’s (2004) results are consistent with the hypothesis that *Australopithecus garhi* belongs to a clade that also includes *Australopithecus africanus*, *Paranthropus* and *Homo*, insofar as *Australopithecus garhi* is reconstructed as the sister taxon of a clade comprising those taxa (Figure 15.4). In addition, the relationships of *Ardipithecus ramidus*, *Australopithecus anamensis* and *Praeanthropus afarensis* are equivalent to those proposed by Asfaw et al. (1999). However, Asfaw et al. (1999) also suggested that *Australopithecus garhi* may be ancestral to all or part of the genus *Homo*. Cladistic analysis fails to find a sister-group relationship between *Homo* and *Australopithecus garhi* (Figure 15.6). Moreover, *Australopithecus garhi* is excluded from a clade that includes only *Homo*, *Paranthropus*, *Australopithecus africanus*, and *K. platyops*. Thus, there is no support for the hypothesis that *Australopithecus garhi* and *Homo* are sister taxa, so *Australopithecus garhi* is unlikely to be the direct ancestor of *Homo*.

### 15.4.3.5 *Orrorin tugenensis*

Found in Late Miocene deposits (Senut et al. 2001), *Orrorin tugenensis* supplanted *Ardipithecus ramidus* as the oldest known fossil hominid. Senut et al. (2001) claim that on the basis of dental and postcranial characters, *O. tugenensis* is the basal member of the *Homo* clade, to the exclusion of australopiths. Moreover, they suggest that *Ardipithecus ramidus* is not a hominid but an ancestor of *Pan*. A cladogram consistent with these hypotheses (Figure 15.5d) would have *Orrorin* and *Homo* as sister taxa, a clade of all australopithecines.
except *Ardipithecus* being the sister taxon of the *Orrorin + Praeanthropus + Homo* clade, and *Ardipithecus* as the sister taxon of *Pan*.

Strait and Grine’s (2004) analysis did not include *Orrorin* because too few characters are preserved in that species. However, their data set can be used to examine the effect of making *Ardipithecus* the sister of *Pan* and the other australopiths monophyletic (Senut et al. 2001). The most parsimonious tree found by Strait and Grine’s data set for the hypothesis that *Ardipithecus* in the sister taxon of *Pan* is 30 steps longer than that shown in Figure 15.6. Considering that Senut et al.’s (2001) hypothesis is based on only a few characters (e.g., molar size, enamel thickness, details of the proximal femur), which cannot account for so many steps, it is fair to conclude that this hypothesis is not favored by cladistic analysis.

### 15.4.3.6 *Kenyanthropus platyops*

The discovery of *Kenyanthropus platyops* was notable because it demonstrated the existence of multiple hominid lineages in the Middle Pliocene. Leakey et al. (2001) noted that *Kenyanthropus platyops* appeared to share several derived character states exclusively with *H. rudolfensis*. They posited (Lieberman 2001) that this might imply that these two species had a particularly close relationship. A cladogram consistent with this hypothesis (Figure 15.5e) would have *K. platyops* and *H. rudolfensis* as sister taxa. Although the validity of the species diagnosis of *Kenyanthropus platyops* has been questioned by White (2003), who has implied that many of the defining features of the type specimen are artifacts of postdepositional distortion, others have found no reason to doubt its validity (see also Volume 3 Chapter 7).

The results of the analysis by Strait and Grine (2004) are inconsistent with the hypothesis that *K. platyops* shares especially close affinities with *H. rudolfensis* even to the point of removing the latter from the *Homo* clade. Rather, *Kenyanthropus* is the sister taxon of either *Paranthropus* or the *Homo + Paranthropus* clade (Figure 15.6). There is no strong evidence supporting the hypothesis that *H. rudolfensis* and *K. platyops* are sister taxa, and thus the transfer of *H. rudolfensis* to the genus *Kenyanthropus* is at present unwarranted. One implication of these results is that some of the facial features shared between *H. rudolfensis* and *K. platyops* may be primitive for the *Homo + Paranthropus* clade, while others may be convergent. Another implication concerns the timing of early hominid cladogenic events. If *K. platyops* is a valid species then its age (3.3–3.5 Ma) and cladistic relationships suggest that *Homo* and *Paranthropus* may have diverged.
from other hominid taxa up to 700 kyr prior to the earliest known specimens currently attributed to those genera (Suwa et al. 1996). It follows, therefore, that this divergence would not be explained by the Turnover Pulse Hypothesis (Vrba 1988) because the divergence would have predated the desiccations event that she postulates to have occurred in Africa between 2.7 and 2.3 Ma. These two clades may have each diversified during this period, but their origins are likely to have been earlier in the fossil record.

### 15.4.3.7 Sahelanthropus tchadensis

The title of “oldest hominid” now belongs to *Sahelanthropus tchadensis* (Brunet et al. 2002). Brunet et al. (2002 p 151) note that *S. tchadensis* appears to be “the oldest and most primitive member of the hominid clade, close to the divergence of hominids and chimpanzees.” The authors are cautious about the precise phylogenetic relationships of the species, but note the possibility that *Sahelanthropus* is the sister taxon of all other hominids, including *Ardipithecus*. A cladogram consistent with this hypothesis would have *Sahelanthropus* as the basal branch of the hominid clade (Figure 15.5f). In contrast, Wolpoff et al. (2002) have suggested instead that *Sahelanthropus* may be more closely allied with *Pan* or *Gorilla* than with hominids.

Brunet et al. (2002) discuss the possibility that *Sahelanthropus* is the sister taxon of all known hominid species, including *Ardipithecus*. The Strait and Grine study is consistent with this hypothesis insofar as *Sahelanthropus* was found to be the basal branch of the hominid clade (Figure 15.6). It also does not group with African apes as suggested by Wolpoff et al. (2002).

### 15.4.3.8 Ardipithecus kadabba

Fossils attributed to *Ardipithecus kadabba* were first assigned to a subspecies of *Ardipithecus ramidus* (Haile-Selassie 2001), but subsequent discoveries led to the elevation of this assemblage to species status (Haile-Selassie et al. 2004). The species is notable for its extremely primitive canine–premolar honing complex. Its describers imply that it is the best candidate to be the sister taxon or ancestor of all other hominids and that fossils of the other two known Miocene species, *O. tugenensis* and *S. tchadensis* are in fact representatives of *Ardipithecus kadabba*. This species was not included in the analysis of Strait and Grine (2004) because it is currently known from only a few body parts.
15.4.4 Phylogenetic relationships within the genus *Homo*

Compared with studies of early hominid evolution in the Late Miocene and Pliocene, research on the phylogeny of Pleistocene hominids (Figure 15.7) is complicated by ongoing debates on the number of species involved. The extremes range from those who, like Mayr in 1950, have argued for a single species in the genus *Homo* (Wolpoff et al. 1994)—thus precluding any phylogeny within the

*Figure 15.7*

Hypothetical cladogram (a) and phylogenetic tree (b) of evolution within the genus *Homo* (modified from Tattersall 1999)
genus—to others who suggest the presence of more than 15 species (Tattersall 1999). However, the majority of researchers recognize, at least for the purposes of discussion, between seven and nine species. *Homo habilis, Homo rudolfensis, Homo ergaster, Homo erectus, Homo heidelbergensis, Homo neanderthalensis,* and *Homo sapiens* are widely recognized, with *Homo antecessor* and *Homo floresiensis* more poorly known and less widely accepted.

As noted above, the relationships of *H. habilis* and *H. rudolfensis* are poorly resolved. Recently on the basis of an assessment of adaptive differences between *Homo* and *Australopithecus,* Wood and Collard (1999a, b) argued that these taxa should be removed from the genus *Homo.* Although their suggestion has yet to be widely adopted. Among other early Pleistocene species, there are ongoing debates over whether *Homo ergaster* from Africa is more closely related to later species of *Homo* than is the mostly Asian *Homo erectus* (e.g., Figure 4b), or whether the two taxa can be distinguished at all (e.g., Wood, 1994; Brauer, 1994; Rightmire, 1992).

Like studies of early *Homo,* studies of phylogenetic relationships among later species of the genus *Homo* are bedeviled by problems of proper taxonomic allocation of fossils to be included in any analysis. Many researchers agree that the descendant of the Early Pleistocene *Homo erectus* (or *Homo ergaster?*) is a Middle Pleistocene taxon usually referred to as *Homo heidelbergensis* (Rightmire, 1998), which in turn may have given rise to both Neandertals and modern humans (Figure 15.7b). However, it has also been suggested that the immediate ancestor of *Homo heidelbergensis* is not *Homo erectus,* but the poorly known *Homo antecessor* from the latest Early Pleistocene of Atapuerca Spain (Bermudez de Castro et al. 2004) and possibly Italy as well (Manzi 2004). The relationship between *Homo heidelbergensis,* *Homo neanderthalensis,* and *Homo sapiens* is also uncertain. Although Neandertals have traditionally been viewed as either a subspecies of *Homo sapiens* or the sister taxon of our species, many authorities now argue that *Homo heidelbergensis* and *Homo neanderthalensis* are sister taxa or even a single anagenetic lineage with no clear break (Arsuaga et al. 1997; Hublin, 1998). From this perspective, *Homo sapiens* is the sister taxon of a *H. heidelbergensis* and *Homo neanderthalensis* clade (Figure 15.7a). In this scheme *Homo sapiens* is the descendant of a distinct early Middle Pleistocene taxon from Africa, usually given the name *Homo rhodesiensis.*

**15.5 Conclusions and future directions**

It is too soon to say whether a consensus will emerge concerning the phylogenetic relationships of the hominid species described over the last decade, not to
mention those likely to be discovered in the coming years. Strait and Grine’s (2004) results on early hominid phylogeny need to be tested by other, independent cladistic analyses, and a rigorous phylogenetic analysis of the genus *Homo* is long overdue. New fossils of almost all of these species are badly needed in order to provide a better representation of characters and a better understanding of intraspecific variation. Improvements in techniques to assess character independence, morphological integration, and developmental modularity (McCullum 1999; Ackermann and Cheverud 2000; Strait, 2001) will also greatly improve the accuracy of cladistic analysis. At the heart of all attempts to understand hominid phylogeny are unresolved issues regarding the identification of species in the fossil record (Tattersall 1986, 1992, 1996; Kimbel and Rak 1993; Plavcan and Cope 2001). This is especially critical within the genus *Homo*.

Despite these caveats, we have a broad consensus regarding the phylogenetic relationships of many hominid taxa (Figure 15.8). The phylogenetic relationships of *Ardipithecus ramidus* and *Australopithecus anamensis* do not appear to be controversial (see also Volume 3 Chapter 7; Kimbel et al. in press). It is likely that disagreement will persist as to the exact relationships among *S. tchadensis*, *Ardipithecus kadabba*, and *O. tugenensis* until they are known by

- **Figure 15.8**  
A summary of the temporal span and phylogenetic relationships among fossil hominids
more body parts that can be directly compared, but most workers accept that these species all lie somewhere near the base of the hominid tree. The greatest disagreement will probably focus on *Australopithecus garhi*, *K. platyops*, and the relationships of these species to the genus *Homo*. While there is broad general agreement about overall phylogenetic relationships in later hominid evolution, there is less consensus about the number of taxa that should be identified. The number and relationships of the early species of the genus *Homo* remain a source of ongoing debate (Wood and Collard 1999a, b) and there are various alternative interpretations concerning the few fossils from the Early and Middle Pleistocene (Tattersall 1986; Rightmire 1998; Bermudez de Castro et al. 2004; Manzi, 2004). Although debates concerning possible interbreeding between Neandertals and modern humans are ongoing and perhaps insoluble, there is a growing consensus that they are morphologically distinct taxa that were separated for hundreds of thousands of years (Krings et al. 1997; Hublin 1998; Stringer 2002).

Some have argued that hominid phylogeny will never be resolved in our time because of the many gaps in the fossil record (White 2002). We find that view unduly pessimistic. There is no doubt that the fossil record samples only a portion of the organisms, including hominids, that have ever lived, and that new discoveries always document new, unanticipated aspects of evolutionary diversity. This is why paleontology is such an exciting and rewarding field of study. Despite this serendipitous sampling of the history of life, and ongoing uncertainties regarding some taxa, cladistic analysis has led us toward a general consensus of the phylogenetic relationship of many of the hominid taxa we can document and remains our best hope for resolving future questions about hominid phylogeny.

**References**


Bräuer G (1982) Early anatomically modern man in Africa and the replacement of the
Mediterranean and European Neanderthals.  


Hooton EA (1931) Up from the Ape. Macmillan, New York


Szalay FS (1977) Ancestors, descendants, sister groups, and testing of phylogenetic hypotheses. Syst Zool 26: 12–18


australopithecines. Aldine de Gruyter, New York, pp 405–426


Abstract

Biomolecules, in particular DNA, assist us in generating and testing hypotheses about human evolutionary history. Molecular analyses testing for and then utilizing a local molecular clock can inform us as to the timing of the split between different lineages or populations. When applied to the split between hominins and chimpanzees, for instance, the molecular clock estimates of their divergence date places constraints on interpretations of the growing fossil record from the Late Miocene and Early Pliocene. The pattern and distribution of modern human variation can be used to extrapolate back in time to infer when and where the modern human gene pool arose. Mitochondrial DNA and Y chromosome sequences and markers have been extensively surveyed in populations from around the world. Numerous nuclear loci and other markers, such as microsatellites and Alu insertions, have similarly been sampled and analyzed. The majority of such analyses point toward a relatively recent origin for modern human diversity from a small population in Africa within the last 200 kyr, with a subsequent dispersal into Eurasia less than 100 ka. The question then remains if these modern migrants met and hybridized with the existing archaic populations. Analyses of ancient mitochondrial sequences from 10 Neanderthals strongly suggest that Neanderthal females did not contribute to the modern human mitochondrial gene pool. Coupled with evidence from the fossil and archeological records, a recent African replacement model best fits the existing data.

16.1 Background

of animals serves to carry us back into geological times, and I believe that we have but begun the work along these lines, and that it will lead to valuable results in the study of various problems of evolution” (Nuttal 1904 p 4).

We now know that biomolecules, in particular DNA, can inform us about phylogeny and population history, selection, and perhaps even taxonomy. Inferences drawn from molecular analyses can provide insights into at least three areas of hominin history. The first is the timing of the hominin–chimpanzee split, which in turn may provide a background for interpreting the growing Late Miocene–Early Pliocene hominin and hominid fossil record. The second is the origins of modern human populations by extrapolating into the past by examining the pattern of modern human molecular variation. Finally, Middle to Late Pleistocene fossils such as the Neanderthals have now successfully yielded DNA sequence information which allows us to draw inferences back to the point at which the modern and Neanderthal lineages originated.

Figure 16.1
Time ranges of hominin species (after Wood 2002). Gray ovals represent the four places where molecular information may be informative. (a) the timing of the split between hominins and chimpanzees, (b) the origins of the modern human gene pool, (c) the diversity and origin of the Neanderthal gene pool, and (d) the time of divergence between Neanderthals and modern humans.
Nuttal’s research, carried out shortly after the discovery of blood groups in 1901, was based on qualitative and quantitative measures of the immunological reactions of various proteins in the blood. Immunological approaches were improved and systematically applied to questions about primate evolutionary history extensively in the 1960s through the works of Goodman (1961, 1963) and Sarich and Wilson (1966, 1967). It was also during this period that the concept of the molecular clock was first proposed (Zuckerkandl and Pauling 1962). By the 1970s, research increased directly at the DNA level, though only using approximate methods such as DNA–DNA hybridization and restriction mapping to measure the differences between species, populations, and individuals. This time period also saw the development of chromosomal banding techniques for evolutionary analysis (Chiarelli 1966; Dutrillaux 1979; Yunis and Prakash 1982). These techniques have been further developed and have helped us understand the rearrangements that are both shared and differ between humans and chimpanzees using fluorescence in situ hybridization (FISH) and reciprocal chromosomal painting (Weinberg and Stanyon 1998). These techniques are generally only used clinically within modern humans as our level of chromosomal variation is extremely low.

Techniques that directly measure differences at the DNA level have advanced greatly in the last two decades. Earlier studies of species and population differences utilized restriction endonucleases, enzymes that cut a strand of DNA at a particular short sequence pattern, to estimate either genetic distances or to provide phylogenetically informative characters between individuals in the sample. Today, the majority of molecular information is derived using variety of high throughput automated technologies that allow DNA sequences, microsatellite allele sizes, and single nucleotide polymorphisms (SNPs) to be rapidly characterized.

While in the ideal world, molecular phylogenetic and population genetic analyses could be carried out using whole genomes, practical constraints require the use of either relatively short sequences (hundreds to tens of thousands of bases) or variable makers such as retroelements, microsatellites, and SNPs. An increasingly popular class of molecular makers are retrotransposable elements, including short interspersed elements (SINEs) and long interspersed elements (LINEs). SINEs, particularly the Alu family, which exists in over 500,000 copies in human genomes, can vary in number and location between individuals and populations (Batzer et al. 1996). Because the absence of an Alu element at a particular location in the genome is the ancestral condition, the shared presence of an element is most likely indicative of common descent. Similarly, the longer
LINE elements, which make up over 15% of human genomes, can be used as markers of common evolutionary descent (Sheen et al. 2000; Boissinot and Furano 2005). Extremely variable short tandem repeats (STRs), also known as microsatellites, have also proven useful in individual identification, parentage assessment, and to infer population relationships based on the analysis of the frequencies of different allele sizes (Bowcock et al. 1994). Finally, SNPs are increasingly being used to infer population relationships and evolutionary history (Yu et al. 2002).

Endogenous retroviruses make up a surprising large portion of the human genome, and their type, copy number, and positions within the genome vary between populations. Therefore, they can provide useful evolutionary markers in the same way as the retrotransposable elements mentioned above (Turner et al. 2001). Extragenomic molecular data from pathogenic and commensal organisms can also be useful in inferring human evolutionary history. Tapeworm, lice, and stomach bacteria sequences have all been used to generate and test hypotheses about human population relationships and migrations (Hoberg et al. 2001; Disotell 2003; Leo and Barker 2005).

By late 2005, a complete draft of the chimpanzee genome was sequenced joining the complete human genome assembly allowing for even more sophisticated comparative analyses (Chimpanzee Sequencing and Analysis Consortium 2005). Not only is the chimpanzee genome useful for investigating the differences (and similarities) between Homo and Pan, it also provides information useful in inferring the character polarity of molecular markers that vary among humans.

Finally, molecular data are being used to investigate the differences between humans and our primate relatives through studies of gene expression and other underlying molecular and developmental process, but these issues are beyond the scope of this review (Enard et al. 2002).

16.3 Hominin origins

Molecular studies have been used to draw inferences about the probably Eurasian origin of the African hominids including the ancestor of hominins, though not without controversy (Miyamoto et al. 1998; Stewart and Disotell 1998; Moyà-Solà et al. 1999). However, until the relatively recent discoveries of the Late Miocene hominids and/or hominins including Ardipithecus, Orrorin, and Sahelanthropus (Haile-Selassie 2001; Senut et al. 2001; Brunet et al. 2002), little could be said about origin of the hominin lineage itself. In fact, only its date, approximately 6 Ma, inferred from molecular clock estimates was available (Chen and Li 2001;
Wildman et al. 2003). This date estimate is not without controversy, though significantly older dates put forth by Arnason et al. (1996, 1998) and supported by Tavare et al. (2002) do not appear to hold up with more detailed molecular analyses (Raum et al. 2005).

With the discovery of these Late Miocene fossils, numerous phylogenetic hypotheses were put forth suggesting their hominin status (Haile-Selassie 2001; Brunet et al. 2002) or that only one was an early hominin and the others were either fossil chimpanzees or gorillas or broadly ancestral to both the human and chimpanzee lineages (Senut et al. 2001). Given the vast amount of molecular data collected to assess the relationships amongst African apes including humans, “… genetic data can also give us trees that are well enough proportioned to be useful to us as paleontologists and that can provide constraints on our ‘flights of fancy,’ when calibrated by plausible paleontological or other historical data” (Pilbeam 1995). An approximately 6-Ma split between humans and chimpanzees, for instance, makes the phylogenetic proposal put forth by Senut et al. (2001), in which Ardipithecus falls along the chimpanzee lineage and Orrorin falls well within the hominin lineage more than 2.5 Ma after a human–chimpanzee divergence (assumed by Senut et al. to have occurred around 8.5 Ma), untenable.

An interesting proposal put forth by Wildman et al. (2003) using a combination of lineage divergence estimates based both on molecular and fossil data would substantially revise the taxonomy of all hominins and our close relatives, the chimpanzees. They propose a time-based phylogenetic classification/taxonomy linking the timing of the origin of a clade to its taxonomic level for most catarrhines. In their scheme, because chimpanzees and humans share a recent common ancestor, only 6 Ma, chimpanzees would be classified within the genus Homo as Homo (Pan) troglodytes and Homo (Pan) pansicus. Consequently, all genera of hominins, including Australopithecus, Paranthropus, and Kenyanthropus would necessarily be sunk into the genus Homo. Depending on their phylogenetic positions and dating, Sahelanthropus, Orrorin, and Ardipithecus might similarly be included within Homo. While the chaos and difficulty of adopting this strategy are apparent, the underlying phylogenetic logic is appealing.

16.3.1 Modern human origins

Most studies of blood group allele frequencies and protein polymorphisms carried out in the 1960s and early 1970s that presented their findings in the form of a phylogenetic tree posited a basal split between Asians and an Afro-European cluster. In 1974, Nei and Roychoudury (1974) analyzed 21 blood group systems and 35 polymorphic proteins from which they inferred an initial African
versus European–Asian split. In this rather prescient chapter, they extrapolated from estimated amino acid replacement rates and inferred that the basal split between Africans and Eurasians occurred approximately 120 ka and that Europeans and Asians split around 55 ka. Few additional studies attempting to infer modern human origins were carried out until the late 1980s.

Two seminal papers published in the late 1980s by Cann et al. (1987) and Vigilant et al. (1989), both working in Allan Wilson’s laboratory at Berkeley, inferred a less than 200-ka African origin for all human mitochondrial DNA (mtDNA) and by extrapolation, perhaps all modern populations. Known by various names, the “Mitochondrial Eve” or “Out-of-Africa” hypothesis, herein it will be referred to as the recent African origin hypothesis. This model stands in contrast to the regional continuity or multiregional model in which local
populations are thought to derive from the original populations who migrated into the various regions of the Old World over 1 Ma from Africa with various amounts of gene flow between the different regions ever since (Wolpoff et al. 2000). Cann et al.’s (1987) study was based on phylogenetic inferences drawn from parsimony analysis of high-resolution restriction mapping of the whole mtDNA genome. To counter criticisms of the precision of restriction mapping, the geographical sampling, and the lack of an outgroup in Cann et al.’s original analysis, Vigilant et al. (1989) employed one of the first PCR analyses of hair samples in a phylogenetic analysis, followed by a sequence-based study with a much larger sample size (Vigilant et al. 1991). Through sequencing the D-loop or control region of the mtDNA genome, they were able to align human sequences with those of a chimpanzee in order to carry out a parsimony analysis rooted with an outgroup. The results were remarkably congruent with those of Cann et al. (1987) in inferring a similar timing and location for the origin of all contemporary human mtDNA, approximately 200 ka in Africa.

The initial papers of Cann et al. (1987) and Vigilant et al. (1989, 1991) came under criticism for an important analytical flaw. The parsimony trees presented suggesting a recent African ancestry for all modern mtDNA were derived from heuristic search strategies that did not find the most parsimonious trees for their respective data sets. Other researchers were able to produce trees without African roots that were more parsimonious (Maddison et al. 1992; Templeton et al. 1992). Since no search strategy is available to guarantee the most parsimonious tree is found for such large data sets, alternative strategies were utilized to infer the root of the modern human mtDNA tree. Stoneking et al. (1992) and Sherry et al. (1994) demonstrated that the much greater amount of mtDNA diversity found within Africa compared to outside of it was best explained by a longer period of time for it to accumulate within Africa. Additional new though smaller data sets chosen to represent the most diverse human sequences possible were also analyzed, in which an African origin for modern mtDNA types were inferred (Kocher and Wilson 1991). Additional molecular dating inferences also supported the approximately 200-kyr time frame inferred to explain human mtDNA diversity (Ruvolo et al. 1994).

More recently, a huge number of human mtDNA sequences have been collected and subject to phylogenetic and population genetic analyses. Due to the rapid rate of evolution of the mtDNA genome, short sequences such as those found in the D-loop or control region are not always useful over long timescales and may show spurious clustering due to homoplasy and multiple substitutions at the same site, including saturation of substitutions at a site. One solution when available is to characterize both the fast-evolving control region and several more slowly evolving region of the mtDNA genome to define haplogroups, or related
lineages of mtDNA haplotypes. In fact, sequencing the complete 16.5-kb mtDNA genome has become commonplace (Ingman et al. 2000; Herrnstadt et al. 2002). By the end of 2005, more than 1,500 complete human mitochondrial genomes have been deposited in GenBank. Such analyses reveal more geographic partitioning of mtDNA sequences than previous studies revealed. Three major haplogroups have been defined for African mtDNA haplotypes, nine for European and seven for Asian (including Native American) mtDNA haplotypes. Most haplogroups have numerous subgroups and some contain a large number of different haplotypes (Herrnstadt et al. 2002). Mitochondrial haplogroup trees contain an African root with Eurasian haplogroups derived from one subgroup of the major African L3 haplogroup (Kivisild et al. 1999; Herrnstadt et al. 2002).

Figure 16.3
Various criticisms of using mtDNA sequence data have been put forth including the possibility of nonmaternal inheritance, selection skewing inferences of geographic structure and rates of evolution, and the presence of recombination. To date, no firm evidence of paternal inheritance has been demonstrated in humans (Bandelt et al. 2005). Furthermore, a mechanism that destroys sperm mtDNA has been discovered making paternal inheritance even more unlikely (Nishimura et al. 2006). Claims for selection acting strongly upon some human mtDNA lineages, especially related human’s entry into colder climates, have been put forth (Mishmar et al. 2003; Ruiz-Pesini et al. 2004). Others interpret the evidence for selection as mainly for purifying selection with only a restricted amount of positive selection in a small portion of the mtDNA genome (Elson et al. 2004). Eyre-Walker and Smith (1999) suggested that mtDNA genomes undergo recombination making inferences about their evolutionary history much less straightforward. The suggestion that mtDNA undergoes recombination has been amply countered by further analyses (Macaulay et al. 1999).

All in all, mtDNA analyses provide a very powerful tool for inferring the evolutionary history of humans and provide a remarkably consistent story as additional data and techniques are brought to bear. Mitochondria, however, only yield a maternal history of the organisms under study. To better understand the overall evolutionary history of any group, both male and biparentally inherited loci are also needed.

The Y chromosome fulfills an analogous paternal role to maternally inherited mtDNA. While it was initially thought that little variation was present on the human Y chromosome, increasingly sophisticated molecular analytical techniques have allowed us to discover a wealth of variation and potential phylogenetically informative markers.

The two major research groups investigating human evolutionary history using Y chromosome markers concur in inferring an relatively recent African origin, around 100 ka or even more recently with a more recent exodus into the rest of the Old World (Hammer et al. 1998; Underhill et al. 2000). This date is more recent than the mtDNA-derived estimate most likely because of the smaller effective population size of the human Y chromosome due to greater variation in reproductive success of males compared with females and the greater geographic structuring of Y chromosome variation. Trees inferred from Y chromosome data generally have their deepest and next deepest roots within African populations (Underhill et al. 2000; Hammer et al. 2001; Jobling and Tyler-Smith 2003).

A study of over 10,000 base pairs on a region of the X chromosome with low levels of recombination also is compatible with the mtDNA and Y chromosome results. Kaessmann et al. (1999) found an approximately 535 ka most recent
common ancestor for this alleles of this region. This is broadly consistent with the mtDNA and Y chromosome dates given that the effective population size of the X chromosome is three times that of the other two loci and therefore coalescent estimates will be approximately three times as old as well.

Nuclear loci such as the compound haplotype composed of an STR locus and an Alu deletion polymorphism on chromosome 12 at the CD4 locus demonstrate a similar pattern to the mtDNA and Y chromosome patterns of variation (Tishkoff et al. 1996). An African origin for the variation at this locus is estimated in the same time frame as that inferred from mtDNA and Y chromosome data with dramatically reduced variation found outside of Africa. Phylogenetic trees derived from numerous microsatellite (STR) loci similarly find their roots within Africa with reduced variation outside of it, though divergence date estimates cannot be easily calculated from such data (Bowcock et al. 1994). A similar pattern is found with SNPs (Yu et al. 2002) polymorphic Alu insertions (Batzer et al. 1996).

Altogether, the majority of analyses of molecular sequences and markers suggest a recent African origin for the diversity of modern human genomes (Jorde et al. 2000; Takahata et al. 2001; Excoffier 2002; Satta and Takahata 2002). However, interpretations that contradict a scenario of a recent African origin have been put forth (Harris and Hey 1999; Hawks and Wolpoff 2001;
Harris and Hey (1999), for instance, interpret PDHA1 (an X chromosome locus) sequence diversity as yielding a 1.86-Ma common ancestor, which would fall outside of the range of estimates derived from the above loci. The PDHA1 analysis has been called into question due to the probability that the locus is under selection making inferences as to coalescence dates difficult (Disotell 1999).

Nested clade analyses (Templeton 2002, 2005) suggest more than one major exodus from Africa, an early one, approximately 1.9 Ma, one around 600–700 ka, and a final one around 100 ka with evidence for range expansion, long-distance dispersal, and isolation by distance complicating the picture. These analyses have generated a healthy skepticism, especially over the efficacy and accuracy of nested clade analysis (Cann 2002; Knowles and Maddison 2002; Satta and Takahata 2002). The important attempt to utilize molecular markers from multiple loci with different effective population sizes and selective pressures in many populations on a global scale is necessarily complicated. As the techniques to analyze such data are further tested and simulated, their potential, power, and limitations will become apparent.

### 16.4 Neanderthals

The final opportunity for biomolecules to shed light on hominin phylogeny involves the direct sequencing of DNA from fossils. The presence of hundreds to thousands of copies of the mtDNA genome in most cells makes it an ideal candidate for extraction from poor or degraded sources of tissue, such as teeth and bone, including fossils. As of the end of 2005, mtDNA sequences from 10 Neanderthal individuals have been gathered (Krings et al. 1997; Ovchinnikov et al. 2000; Schmitz et al. 2002; Serre et al. 2004; Beauval et al. 2005; Lalueza-Fox et al. 2005). These sequences form a reciprocally monophyletic clade with the thousands of modern human mtDNA sequences analyzed to date.

Ancient DNA (aDNA) analyses are however fraught with difficulties (Cooper and Poinar 2000; Mulligan 2005). Ancient DNA, when present, even under ideal preservation conditions is likely to be damaged and fragmented. More importantly, it is almost certainly contaminated with modern DNA from the environment, excavators, curators, scientists who have handled the material, and molecular laboratory personnel. Extraordinary precautions and techniques need to be carried out to lower the probability of mistakenly accepting such modern contaminants as the sequences from the ancient material (Cooper and Poinar 1998, 2001). Despite these difficulties, aDNA provides a unique and important window in the evolutionary history and processes.
Given the current sampling of 10 individuals from different time periods and geographic locations, it is unlikely that a Neanderthal sequence that falls within the modern mtDNA gene pool will be discovered (Krings et al. 2000). Wolpoff (1998) suggested that because the original Feldhofer Neanderthal sequence is more similar to some modern human sequences than some other modern sequences are to other moderns, their mtDNA gene pools overlapped. This however was a misleading analysis as a cladistic analysis of the same data clearly demonstrates a complete separation of Neanderthals and moderns into reciprocally monophyletic clades (Disotell 1999). This observation has been further strengthened by all additional Neanderthal sequences and analyses.

Molecular clock estimates suggest the Neanderthal clade last shared a common ancestor with the modern human clade approximately 600 ka (Krings et al. 1997; Ovchinnikov et al. 2000). Interestingly, the genetic diversity among the Neandethals approximates that of contemporary humans, suggesting that they, like modern humans, probably rapidly expanded after passing through a bottleneck. The Neanderthal sequences do not cluster among modern European sequences as might be expected if they gave rise to the Europeans or extensively interbred with the new migrants into Europe as would be predicted under the multiregional model. However, both Nordborg (1998) and Relethford (2001) point out that different amounts of crossbreeding between Neanderthals and early moderns could have still been possible with the Neanderthal mtDNA lineages having gone extinct due to normal stochastic processes over the last 30+ kyr.

To further test hypotheses of modern human origins, several researchers have attempted to recover and sequence early modern human aDNA. One of these attempts, to sequence the Lake Mungo 3 specimen from Australia, provides a good case and point of the numerous difficulties of aDNA analysis. Adcock et al. (2001) claimed to have recovered an mtDNA sequence from an early modern human fossil skeleton from Australia, known as Lake Mungo III, then thought to date to approximately 60 ka [this specimen has since been redated to 40 ka (Bowler et al. 2003)]. The sequence fell outside of the range of modern human mtDNA diversity, and clustered with a sequence located on chromosome 11 of the modern human genome, a known mitochondrial pseudogene (numt). Their interpretation was that early modern humans reached Australia before the most recent African exodus that gave rise to the rest of the world’s mtDNA diversity less than 100 ka.

This analysis seems deeply flawed for several reasons. First, the standard protocols suggested to avoid contamination with modern DNA (Cooper and Poinar 1998; Mulligan 2005) were not rigidly followed (Cooper et al. 2001). The sequence is most likely in fact a contaminating numt or has been damaged to
yield spurious nucleotide substitutions (Cooper et al. 2001). Smith et al. (2003) point out that it is extremely unlikely for aDNA to have survived at the Lake Mungo site due to the environmental conditions present. Finally, reanalysis with additional Australian and African sequences yields a tree very different from that originally put forth (Cooper et al. 2001). Caramelli et al. (2003) attempted to sequence several early modern specimens from Paglicci cave in Southern Italy. Their sequences fully fall within the range of modern human sequences. These sequences are therefore either modern contaminants, or, early modern mtDNA sequences indeed fall within the range of all modern mtDNA present today. Until we can train chimpanzees (whose contaminating DNA will be easy to identify) to excavate fossils, extract DNA, and carry out laboratory analyses, it seems unlikely that we will be able to definitively identify DNA recovered from early modern samples.

Serre et al. (2004) therefore took a different approach to investigating early modern human and Neanderthal mitochondrial diversity. They realized that demonstrating the presence of early modern mtDNA is nearly impossible, so they tested five early modern fossil samples along with four Neanderthal samples for the presence of Neanderthal-specific mtDNA motifs. Included among the early human samples were fossils from Vindija, Croatia, and Mladec (Czech Republic) that have been claimed to be transitional between Neanderthals and early moderns (Wolpoff 1999). Their reasoning was that if interbreeding occurred between the two groups, the presence of Neanderthal mtDNA in early modern individuals would be more likely since it would not have had a great amount of time to go extinct as Nordborg (1998) and Relethford (2001) potentially proposed for the absence of Neanderthal mtDNA today. Serre et al. (2004) were able to amplify all four Neanderthal samples with "Neanderthal-specific" primers. None of the early modern human fossils yielded amplification products though they did for more generalized "hominoid-specific" primers suggesting DNA was present. Furthermore, faunal samples from the same sites all yielded DNA products suggesting that the conditions at the sites were adequate for the preservation of aDNA.

While many more early modern samples would have to be tested to statistically rule out a Neanderthal contribution to the modern mtDNA gene pools, the sum total of all the evidence to date makes it extremely unlikely that such contributions will be found. Given the difficulty of retrieving and sequencing bona fide aDNA, it seems unlikely that we will be successful in the near future of finding fossil Y chromosome or nuclear sequences. The only inferences we can draw about Neanderthal contributions to these portions of the genome are from inferences from modern variation. If indeed, Neanderthals and human split
around 0.5 Ma or earlier, the depth of modern Y chromosome diversity suggests that like the case for mtDNA, there is little to no Neanderthal contributions present today.

16.5 Conclusions

Biomolecules have many advantages over morphological characters for phylogenetic analyses. The sheer volume of data potentially available is staggering. More importantly, hypotheses regarding homology, which not trivial, are generally more robust than those inferred for morphological characters and systems. The independence of characters and traits is more easily achieved at the molecular level allowing multiple independent phylogenetic hypotheses to be generated and examined for concordance. On the other hand, all molecular phylogenies are necessarily gene trees which can have a different history from the species or populations in which they reside. This necessitates the study of multiple independent loci whenever possible. Homoplasy and selection are more easily detectable at the molecular level. Fossils on the other hand can test hypotheses that have been put forth and suggest novel combinations of traits that we are not clever enough to have thought possible. In order to better understand hominin evolution, a combination of approaches and techniques will provide us with the best insights into our evolutionary history.

References


Phylogenetic relationships (biomolecules)
Nei M, Roychoudury AK (1974) Genic variation within and between the three major
Nuttal GHF (1904) Blood immunity and blood relationships. Cambridge University Press, Cambridge
17 Population Biology and Population Genetics of Pleistocene Hominins

Alan R. Templeton

Abstract

The population genetics of Pleistocene hominins is deduced from three types of data: coalescent processes and haplotype trees estimated from surveys of genetic variation in present-day human populations, haplotypes inferred from ancient DNA extracted from fossils, and overlays of current quantitative genetic variance/covariance matrices upon hominin fossils. The haplotype trees are subjected to nested clade phylogeographic analyses, including many new analyses never before published. These analyses show that there were three major expansion events of hominins out of Africa during the last 2 million years. The first expansion event marked the original dispersal of Homo erectus out of Africa into Eurasia. The quantitative genetic analysis of hominin fossils indicates that there was relaxed selection upon at least some morphological features at this time, perhaps due to an increased use of cultural inheritance in dealing with the environment. Coalescent analyses indicate that the colonization of Eurasia was marked by strong selection at many loci, so although morphological selection may have been relaxed, adaptive processes were still proceeding as humans colonized this new geographical area. Eurasian and African populations also established recurrent gene flow restricted by isolation by distance by 1.46 million years ago. A second expansion out of Africa was marked by the spread of the Acheulean culture, implying that the spread of this culture was due to a spread of peoples and not just ideas. The expanding Acheulean populations interbred with existing Eurasian populations, and recurrent gene flow continued after the Acheulean expansion. A third expansion out of Africa marked the spread of many anatomically modern traits that had earlier appeared in Africa. This expansion was also marked by interbreeding, so regional continuity persisted for some traits. Although total replacement of Eurasian populations is rejected with a $p < 10^{-17}$, it is still possible that some local populations were replaced. Ancient DNA studies are inconclusive about the status of Neanderthals in this regard. Coalescent studies are also inconclusive and contradictory about the size of hominin populations before
this last out-of-Africa expansion and the degree of population growth during the expansion phase. Because of interbreeding and gene flow, humanity evolved into its modern form as a single evolutionary lineage but with some geographical differentiation at any given time.

17.1 Introduction

Population genetics is concerned with the origin, amount and distribution of genetic variability present in populations of organisms, and the fate of this variability through space and time. Variation in genes through space and time constitute the fundamental basis of evolutionary change, so population genetics can be thought of as the science of the mechanisms responsible for evolution within a species or within a continuous lineage of species through time. The fate of genes through space and time is strongly influenced by a population’s demography, so genetic studies can also be used as a powerful tool to investigate demography.

Most population genetic studies of natural populations involve surveying a sample of individuals from one or more demes (local subpopulations), followed by analyses to infer population structure, demography, and/or the impact of various evolutionary forces, such as genetic drift, gene flow, and natural selection. But how does one study the genetics and demography of populations from the distant past, such as hominin populations from the Pleistocene? The most common approach to the study of past populations arises from the subdiscipline of population genetics known as coalescent theory. Because DNA can replicate and pass on copies of itself to the next generation, contemporary DNA contains information from past generations. One can therefore genetically survey contemporary populations and then use phylogenetic techniques to estimate the pattern of past DNA replication events. A DNA replication event produces two copies of DNA from one ancestral molecule, so when looked at backward in time, a DNA replication event corresponds to two molecules of DNA coalescing into a single ancestral molecule. Our ability to infer coalescent events with phylogenetic methods depends on the descendant DNA lineages being distinguishable from one another, so the only coalescent events that can be inferred are those that are also associated with a mutation in one of the DNA lineages. The distinguishable DNA lineages are called haplotypes, and therefore we can only reconstruct the haplotype tree of a DNA region; that is, the pattern of coalescent events also marked by mutational changes. Thus, the rate of mutation in a DNA region places a limit on our ability to infer the evolutionary history of that region. Because of the Mendelian properties of genetic recombination and assortment
and because different DNA regions can display different patterns of inheritance (e.g., the DNA on the Y-chromosome is paternally inherited but autosomal DNA is inherited through both sexes), different regions of DNA can have different evolutionary histories and can be influenced by different subsets of past demographic events and processes. Moreover, population genetic inference requires genetic variation, so once all the contemporary copies of a gene or DNA region have coalesced back to a single ancestral molecule of DNA, all population genetic information is lost. Because different DNA regions can coalesce to their common ancestral molecule at different times in the past, different genes are informative of different time periods in the past. For all of these reasons, any one gene or DNA region captures only a portion of a population's evolutionary history, and the haplotype tree of a gene or DNA region should never be equated to the evolutionary history of the population. Many different haplotype trees are required for a full reconstruction of population history. Fortunately, such multilocus, coalescent analyses are now feasible.

One limitation of the coalescent approach is not easily overcome. Coalescent analyses are based on the evolutionary history of current genetic variation, and therefore are limited to inferences on past populations or subpopulations that have left genetic descendants in present-day populations. Any population that left no descendants is outside the domain of coalescent analyses. DNA from fossils can overcome this limitation, so ancient DNA studies are complementary to coalescent analysis. At present, the number of fossils that retain useful DNA severely limits ancient DNA approaches. As a result, ancient DNA surveys usually involve limited sample sizes, even when fossils from diverse geographical locations and separated by tens of thousands of years are pooled together as a single “population.” Extremely small sample sizes preclude precise population genetic inference. Moreover, ancient DNA studies on hominins are currently limited to mitochondrial DNA (mtDNA), which is maternally inherited. This limitation guarantees an incomplete inference that can rarely be biologically conclusive.

Coalescent approaches can also be used to infer the presence and type of natural selection operating upon specific genes. Hence, some aspects of adaptive evolution in hominin populations can be addressed from surveys of present-day genetic variation. Another approach to studying past adaptive evolution is to use genetic studies on morphological variation in present-day populations to model the evolutionary forces that created morphological change in the fossil record. All three approaches—coalescence, ancient DNA, and quantitative genetics—will be considered in this chapter with respect to two major aspects of hominin evolutionary history: (a) hominin population structure and historical demographic events and (b) natural selection and adaptive evolution in past hominin populations.
17.2 Hominin population structure and historical demographic events

17.2.1 Nested clade phylogeographic analyses

Nested clade phylogeographic analysis (NCPA) is a coalescent-based approach to extract information from haplotype trees to infer the qualitative nature of past population structure (patterns of gene flow among geographical areas) and historical demographic events (population range expansions, colonization events, and fragmentation events) (Templeton et al. 1995). NCPA first defines a series of hierarchically nested clades (branches within branches) from the haplotype tree using a set of explicit nesting rules (Templeton et al. 1987; Templeton and Sing 1993). Most human haplotype trees are rooted, so the oldest clade is known in any given nested category. The relative temporal orderings are used to analyze the spread of haplotypes and clades through space and time. NCPA next quantifies the spatial distribution of haplotypes and clades by measuring how widespread a clade is spatially and how far away a clade is located from those clades with which it is nested into a higher level clade (Templeton et al. 1995). To adjust for sampling, the nested clade analysis uses a random permutation procedure to test the null hypothesis that the clades nested within a higher level clade have no geographical associations. All subsequent inferences are limited to those clades associated with a statistically significant rejection of the null hypothesis of no geographical association.

Statistical significance tells us that geographical associations exist within the haplotype tree, but they do not tell us how to interpret those geographical associations. Indeed, no single test statistic discriminates between recurrent gene flow, past fragmentation, and past range expansion in NCPA. Rather, it is a pattern formed from several statistics that allows discrimination. Also, many different patterns can sometimes lead to the same biological conclusion, and sometimes a statistically significant pattern has no clear biological meaning because of inadequate geographical sampling or a lack of genetic resolution. Finally, NCPA searches out multiple, overlaying patterns within the same data set. In light of these complexities (which reflect the reality of evolutionary possibilities and sampling constraints), an inference key was provided as an appendix to Templeton et al. (1995), with the latest version being available at http://darwin.uvigo.es/ along with the program GEODIS for implementing the nested clade analysis.

Although NCPA has many strengths, it does have limitations. In particular, inference is limited by (a) sample size and sample sites, (b) insufficient genetic resolution to detect an event or process that actually occurred, and (c) false
inferences arising from the evolutionary stochasticity of the coalescent process or by the haplotype tree being skewed or otherwise altered by natural selection. In light of these limitations, the inference key has been extensively validated by applying NCPA to actual data sets with 150 \textit{a priori} expectations (Templeton 2004b). The inference key did well, with the most common error being the failure to detect an expected event. Only rarely did NCPA result in a false positive. The failure to detect known events was due to the fact that an appropriate mutation had not occurred in the right place and time to mark the event. This shows that no one locus or DNA region can capture the totality of a species’ population structure and evolutionary history. Concerning the rare false positives, the processes of mutation and genetic drift, which shape the haplotype tree upon which the NCPA is based, are both random processes. Therefore, the expected pattern for a particular event or process can sometimes arise by chance alone, leading to a false biological inference. Moreover, natural selection can lead to false biological inferences by skewing the shape of the haplotype tree and the geographical distribution of certain haplotypes.

The occurrence of false negatives and false positives can be reduced by performing NCPA upon many loci or gene regions (Templeton 2002). Using multiple DNA regions reduces the danger of missing an event or process due to the lack of an appropriately placed mutation in time and space for any one DNA region. The chance of making a false inference is reduced by cross-validating inferences across DNA regions. Templeton (2002) used a multilocus NCPA based on the human mitochondrial genome and nine nuclear genome regions to infer recent human evolutionary history. What was most remarkable about the cross-validated inferences in this case was the incompleteness found in any one DNA region. This illustrates that failure to detect events or processes with a single DNA region is a common phenomenon. Interestingly, most inferences that were made with one gene were cross-validated by other DNA regions, thereby indicating that the false inferences are rare in NCPA. This is the same pattern observed when validating the inference key with real data with \textit{a priori} expectations (Templeton 2004b).

Cross-validation in Templeton (2002) was based on temporal concordance across DNA regions of NCPA inferences of the same type and geographical location. Assessing concordance of inference type and locality is straightforward as these are categorical variables, but estimated times of events are quantitative variables with considerable error due to the high stochasticity associated with the coalescent process (Templeton 2004b). Therefore, Templeton (2004a) developed a log-likelihood ratio test of temporal concordance. These likelihood ratio tests confirmed the conclusion in Templeton (2002) that there were two separate out-of-Africa expansion events subsequent to the original out-of-Africa expansion.
of H. erectus populations at least 1.7 Ma (million years ago). These tests also strongly rejected the replacement hypothesis for the most recent out-of-Africa expansion event (Templeton 2004b). This maximum-likelihood framework also placed a 95% confidence limit of 610,000 years ago on how long recurrent gene flow has occurred between Africans and Eurasians.

Many additional DNA data sets have become available since the publication of Templeton (2002). Accordingly, the analyses of Templeton (2002, 2004a) are now repeated with an expanded data set consisting of the original 10 data sets and 15 additional nuclear gene regions: the X-linked AMELX locus (Hammer et al. 2004), the X-linked APLX locus (Hammer et al. 2004), the autosomal CCR5 locus (Bamshad et al. 2002), the autosomal CYP1A2 locus (Wooding et al. 2002), the X-linked FIX locus (Harris and Hey 2001), the autosomal FUT2 locus (Koda et al. 2001), the autosomal FUT6 locus (Koda et al. 2001), the X-linked G6PD locus (Saunders et al. 2002; Verrelli et al. 2002), the autosomal HFE locus (Toomajian and Kreitman 2002), the X-linked HS571B2 locus (Yu et al. 2002), the autosomal LACTASE locus (Hollox et al. 2001), the X-linked MAO genes (Balcuniene et al. 2001), the X-linked MSN/ALAS2 gene region (Nachman et al. 2004), the X-linked RRM2P4 locus (Hammer et al. 2004; Garrigan et al. 2005), and the X-linked TNFSF5 locus (Hammer et al. 2004). These new loci and DNA regions were analyzed in the same manner as the original ten, as described in Templeton (2002, 2004a). Because the output of even a single NCPA is quite lengthy, the full analyses are not given here, but are available upon request to the author. What follows is a summary of the most relevant results of the analyses.

Coalescent theory predicts that there will be a large variance in coalescent times to the most recent common ancestor (TMRCA) among the various gene regions, with the unisexual, haploid regions (mtDNA and Y-DNA) having an expected coalescence time of $N_{ef}$ generations (the long-term inbreeding effective size of humanity), the X-linked loci having an expected coalescence time of $3N_{ef}$, and the autosomal loci having an expected coalescence time of $4N_{ef}$ (Templeton 2002). Table 17.1 gives the TMRCAs for the 10 loci used in the study by Templeton (2002) and for the 15 loci added in this analysis. The average coalescence times do indeed fit the expected ranking, and there is also the large expected variance within each category. Table 17.1 further shows that there is a rather continuous temporal coverage up to about 2.5 Ma. Thus, with these additional loci, events going back to around 2 Ma can be inferred with potential cross-validation, which represents a significant improvement in the informative temporal range over that of the original 10 loci (Templeton 2002).

The NCPA of AMELX, APLX, and MAO either resulted in no significant deviations from the null hypothesis of no geographical associations or had only biologically uninterpretable patterns because of inadequate geographical
sampling or insufficient genetic resolution. All other loci had significant and biologically interpretable patterns (although many included some uninterpretable patterns as well). Table 17.2 shows the inferences made from the 15 new loci along with their estimated times of occurrence. Some events could not be timed because there were no mutations available to obtain a reliable estimate. The inferences shown in Table 17.2 overlay in a completely concordant fashion upon the earlier inferences made on 10 loci (Templeton 2002). However, this expanded data set greatly extends our knowledge about human evolution in the 1–2 Ma interval.

Table 17.2 and the previous analyses revealed 15 different identifications of out-of-Africa range expansion events or Africa/Eurasian range expansion events of ambiguous origin. These 15 events are concordant by type (range expansion) and geographical location (expansion events involving both African and Eurasian populations, and when geographical resolution of the origin is possible, it is always out of Africa). To test temporal concordance of these 15 inferences, their
Table 17.2

The NCPA inferences of range expansion events (middle column) and of restricted gene flow between Africa and Eurasia (right-hand column) and their estimated times (in Ma) from the 12 newly analyzed loci that yielded significant and interpretable results

<table>
<thead>
<tr>
<th>Locus</th>
<th>Event Description</th>
<th>Estimated Time (Ma)</th>
<th>Methodology</th>
<th>Estimated Time (Ma)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCR5</td>
<td>Isolation-by-distance</td>
<td>t = 1.9844</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP1A2</td>
<td>Out-of-Africa, t = 1.43</td>
<td>Isolation-by-distance, t = 0.2649</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 0.3532</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FIX</td>
<td>Out-of-Africa, t = 2.686</td>
<td>With some long-distance dispersal, recent</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FUT2</td>
<td></td>
<td>Isolation-by-distance, t = 0.9948</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FUT6</td>
<td></td>
<td>Isolation-by-distance, t = 1.5917</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G6PD</td>
<td>To the Americas, recent</td>
<td>Isolation-by-distance, t = 0.4808</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 0.169</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 0.2535</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 0.5493</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 0.1558</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 0.86</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, recent</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 1.93</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 0.625</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 0.2649</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 0.3532</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 0.09948</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 1.5917</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 1.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 1.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 1.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 1.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 0.4808</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 0.169</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 0.2535</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 0.5493</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 0.1558</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 0.86</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, recent</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 1.93</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 0.625</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 0.2649</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 0.3532</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 0.09948</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 1.5917</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 1.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 1.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 0.4808</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 0.169</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 0.2535</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 0.5493</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 0.1558</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 0.86</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, recent</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 1.93</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 0.625</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 0.2649</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 0.3532</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 0.09948</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 1.5917</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 1.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 1.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 0.4808</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 0.169</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 0.2535</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 0.5493</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 0.1558</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 0.86</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, recent</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 1.93</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 0.625</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 0.2649</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 0.3532</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 0.09948</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 1.5917</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 1.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 1.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 0.4808</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 0.169</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 0.2535</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 0.5493</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 0.1558</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 0.86</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, recent</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 1.93</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 0.625</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Isolation-by-distance, t = 0.2649</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Time is indicated only as “recent” for events too young to be reliably dated by phylogenetic means.

Estimated times are regarded as random variables with a mean given by a standard phylogenetic estimator (Takahata et al. 2001) of the age of the youngest clade contributing in a statistically significant fashion to the inference, with a calibration point of 6 Ma for the divergence between humans and chimpanzees (Templeton 2004a). The variance of this time is given by (Tajima 1983)

$$\sigma^2 = \frac{T^2}{(1 + k)}$$  \hspace{1cm} (17.1)

where $T$ is the standard phylogenetic estimator of age and $k$ the pairwise divergence among present-day haplotypes as measured by the number of mutations that have accumulated in the descendants of the haplotype or node whose age is estimated to be $T$. Equation (17.1) incorporates two sources of error into the variance associated with estimator $T$. First, the numerator of Eq. (17.1) is $T^2$, and
reflecting the evolutionary stochasticity of the coalescent process itself in which the variance is proportional to the square of the mean (Hudson 1990; Donnelly and Tavare 1995). The other factor that influences the variance is \( k \), which depends on the number of mutations that have accumulated in the DNA lineages from \( T \) to the present. This can vary considerably from locus to locus, depending on the local substitution rates and upon the amount of DNA being sequenced. Because \( k \) is generally very small for recent events, phylogenetic dating procedures are often unreliable for recent events (Rannala and Bertorelle 2001).

If the population under study has not been significantly fragmented into subpopulations and evolution is neutral, then the distribution of \( t_i \), the time of a phylogeographic event or process inferred from DNA region \( i \), can be approximated by a gamma distribution (Kimura 1970):

\[
f(t_i|T_i, k_i) = \frac{t_i^{k_i} e^{-t_i/(1+k_i)/T_i}}{(T_i/t_i)^{1+k_i} \Gamma(1+k_i)}
\]

where \( k_i \) is the average pairwise nucleotide diversity among the haplotypes in DNA region \( i \) in the youngest monophyletic clade that contributed in a statistically significant fashion to the NCPA inference of interest and \( T_i \) is the age obtained by the phylogenetic estimator (Templeton 2004a). Given that the NCPA did not reveal any cross-validated fragmentation events, and none of any sort when inference is limited to Africa and Eurasia (the exclusive focus of this chapter), the gamma distribution assumption is justified for the human data.

Templeton (2004a) used these gamma distributions to derive maximum-likelihood estimators of the time of an event based on multilocus data and to derive a log-likelihood ratio test of the null hypothesis that \( n \) separate inferences of a geographically concordant event are also temporally concordant; that is, they are the same event. Figure 17.1 shows the gamma distributions for the 15 inferences of range expansion involving African and Eurasian populations. The log-likelihood ratio test rejects the null hypothesis that all 15 events are temporally concordant with a probability value of \( 3.89 \times 10^{-15} \). Thus, the genetic evidence is overwhelming that there were multiple range expansion events out of Africa during the last 2 Myr (million years) of human evolution. An inspection of Figure 17.1 reveals that the time distributions for the 15 events cluster into three distinct groupings. Accordingly, the null hypotheses of temporal concordance within each of these three groupings were tested, and in all cases there was strong concordance within \( p = 0.95 \) for the most recent expansion out-of-Africa, \( p = 0.51 \) for the middle expansion, and \( p = 0.62 \) for the oldest expansion). Pooling together the inferences from \( j \) homogeneous loci also results in a gamma distribution with mean and variance (Templeton 2004a):
Combining Eq. 17.1 with Eq. 17.4, the effective number of informative mutations about the age of the event based on pooled data, $k_{\text{eff}}$, is given by:

$$k_{\text{eff}} = \frac{\hat{T}^2}{\text{Var}(\hat{T})} - 1 \quad (17.5)$$

The standard log-likelihood ratio test given in Templeton (2004a) can be used to test the null hypothesis that two or more times based on pooled data are the same by using Eqs. 17.3 and 17.5. The log-likelihood ratio test of temporal
homogeneity of the most recent and middle out-of-Africa expansion events yields a chi-square statistic of 40.84 with 1 degree of freedom with a $p$ value of $1.66 \times 10^{-10}$. Hence, the null hypothesis of temporal concordance is strongly rejected, and the first two clusters shown in Figure 17.1 define two distinct out-of-Africa expansion events. The log-likelihood ratio test of homogeneity of the middle and oldest out-of-Africa expansion events also rejects temporal concordance with a chi-square statistic of 8.85 with 1 degree of freedom and a $p$ value of 0.0029. Hence, all three clusters shown in Figure 17.1 identify separate events that are all cross-validated by multiple loci. The estimated times and 95% confidence intervals for these three out-of-Africa range expansions are shown in Table 17.3. The earlier analysis of 10 loci (Templeton 2002, 2004a) detected only the two most recent expansion events. The expanded analysis not only strongly confirms these first two out-of-Africa expansions with new loci but in addition detects a third, older expansion out of Africa.

These NCPA inferences from molecular genetic data are consistent with the fossil and archeological record. The oldest expansion, dated to 1.9 Ma, corresponds well to the fossil dating of the original expansion of *H. erectus* (which includes *ergaster* in this chapter) from Africa into southern Eurasia (Aguirre and Carbonell 2001; Bar-Yosef and Belfer-Cohen 2001; Antón et al. 2002; Vekua et al. 2002). Dennel (2003) has argued that these Early Pleistocene fossil finds in Eurasia may not have represented permanent settlement, but also acknowledges the difficulty of inferring (or rejecting) regional continuity over a long time period from temporally sporadic fossil and archeological finds from a single geographical area. Genetics can offer an important tool that complements the fossil and archeological data. Recall that the NCPA can only detect the genetic signatures of past populations that have made at least some genetic contribution to current populations. This statistically significant signal of an expansion into Eurasia in the Late Pliocene to Early Pleistocene that is cross-validated by three genes would not exist at all if these initial Eurasian colonies had not left some

<table>
<thead>
<tr>
<th>Time up Expansion (Ma)</th>
<th>More Recent Age Limit</th>
<th>Older Age Limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1304</td>
<td>0.0965</td>
<td>0.1693</td>
</tr>
<tr>
<td>0.6508</td>
<td>0.3917</td>
<td>0.9745</td>
</tr>
<tr>
<td>1.9007</td>
<td>0.9937</td>
<td>3.0969</td>
</tr>
</tbody>
</table>
descendants in present-day Eurasian populations. Hence, the initial colonization of Eurasia by *H. erectus* was a successful one that resulted in permanent settlement of at least parts of Eurasia.

The middle expansion out-of-Africa shown in Table 17.3 is consistent with the spread of the Acheulean stone tool culture out of Africa and into Eurasia. Evidence for this culture is found first in Africa at 1.5 Ma (Asfaw et al. 1992), with the earliest non-African site found in Israel about 1.4 Ma. However, Acheulean sites are not widespread in Eurasia until about 0.6–0.8 Ma, particularly eastern Asia (Hou et al. 2000). This has lead to the suggestion by some that there were two Acheulean expansions out of Africa; the first at about 1.4 Ma and the second at 0.6–0.8 Ma (Goren-Inbar et al. 2000; Aguirre and Carbonell 2001; Bar-Yosef and Belfer-Cohen 2001). As outlined above, the current analysis indicates a statistically significant out-of-Africa expansion at 0.65 Ma (0.3917–0.9745 Ma), which corresponds well with the second, more widespread, Acheulean expansion. However, this genetic analysis does not falsify the hypothesis that there was an earlier Acheulean expansion at 1.4 Ma. Indeed, the out-of-Africa expansion detected by *CYP2A* dates to 1.43 Ma. However, because of the large variances associated with older coalescent-based estimates of age (Eq. 17.1), this event at 1.43 Ma could not be distinguished from the older out-of-Africa events detected by *FUT2* and *Lactase* with the likelihood ratio tests described above. Thus, there may well have been an Acheulean expansion at 1.4 Ma, but because this event is between the original out-of-Africa expansion by *H. erectus* and the much stronger (in terms of both archeological and genetic evidence) Acheulean expansion between 0.6 and 0.8 Ma, it will take a much larger data set to achieve statistical discrimination. Another explanation is that the Acheulean expansion at 1.4 Ma may have been very limited geographically, perhaps confined to the Middle East. A limited range expansion from Africa into the Middle East would not be detectable by most of the loci analyzed here because of sparse geographical sampling in Eurasia that often did not include the Middle East. Finally, it is also possible that the 1.4-Ma Acheulean expansion was only a cultural expansion, in which case it would be invisible to NCPA.

It is not clear from the archeological evidence whether the 0.6- to 0.8-Acheulean expansion was solely the diffusion of ideas from Africa or also involved movement of populations or individuals (Saragusti and Goren-Inbar 2001). Here, the genetic data can complement the archeological data. NCPA can only detect movements of reproducing populations and individuals, not ideas. By combining NCPA with archeology, it is likely that the 0.6- to 0.8-Acheulean expansion represented a movement of both people and ideas. Another question that cannot be answered by the archeological data alone is what happened when these Acheulean peoples coming out of Africa encountered the Eurasian
populations? Perhaps the Acheulean peoples drove the earlier inhabitants to extinction, completely replacing them. Alternatively, the expanding Acheulean peoples could have interbred with the Eurasian populations. The Acheulean replacement hypothesis can be tested by noting that if complete replacement had occurred, there would be no genetic signatures of events or genetic processes in Eurasia that would be older than this expansion event (Templeton 2004a). This prediction stems from the simple fact that NCPA can only detect events and processes that affected past populations that left genetic descendants in present-day populations. The Acheulean replacement hypothesis can therefore be tested by testing the null hypothesis that the gamma distribution marking the Acheulean expansion has a mean time that is not significantly older than other Eurasian events or processes with older estimated times. To be conservative in the definition of “older,” an event or processes will only be regarded as older than the Acheulean expansion if its estimated age falls in the older 1% tail of the pooled gamma distribution that describes the Acheulean expansion. This 1% cutoff is calculated from the pooled Acheulean gamma distribution to be 1.0476 Ma. NCPA identified five events or processes with estimated times older than 1.0476 Ma; the first out-of-Africa expansion at 1.9 Ma and four inferences of restricted gene flow dating from 1.25 to 3.4 Ma (Templeton 2002; Table 17.2). The log-likelihood ratio test of the null hypothesis that these five other events and processes involving Eurasian populations are no older than the Acheulean expansion yields a chi-square of 13.66 with 5 degrees of freedom and \( p = 0.0179 \). As described in Templeton (2002), the oldest of these inferences (restricted gene flow with isolation by distance at 3.4 Ma) is based on the MX1 gene that is an outlier in many ways, including having a coalescent time older than the human–chimpanzee split (a phenomenon known as trans-specific polymorphism). Accordingly, Templeton (2002, 2004a) repeated all analyses excluding this outlier gene in order to be conservative. Doing so now reduces the chi-square to 10.37 with 4 degrees of freedom, which is still significant at the 5% level (\( p = 0.0346 \)). Hence, regardless of whether MX1 is included, the null hypothesis of Acheulean replacement is rejected. The expansion of people from Africa about 0.6–0.8 Ma was therefore marked both by bringing a new culture to and by interbreeding with Eurasian populations.

The range expansion from Africa dated to 0.65 Ma (Table 17.3) is also compatible with the fossil record. After the initial expansion of *H. erectus* out of Africa about 1.9 Ma (Table 17.3), there was little change in average brain size up to 0.7 to 0.6 Ma, after which cranial capacities show a substantial increase (Ruff et al. 1997; Relethford 2001b; Rightmire 2004). Hence, the fossil record, the archeological record, and the genetic analysis presented here all imply that an important transition in human evolution occurred about 0.65 Ma.
The final out-of-Africa expansion is dated to 96,000–169,000 years ago and corresponds to a time period in the fossil record when many anatomically modern traits began to expand out of Africa after appearing earlier in African fossils (Stringer 2002; White et al. 2003). There has been considerable controversy over whether this most recent out-of-Africa expansion event was also a replacement event. This recent replacement hypothesis can be tested in the same manner as the Acheulean replacement hypothesis. The older 1% tail of the gamma distribution that describes this recent out-of-Africa expansion occurs at 0.1774 Ma. Two older out-of-Africa expansion events occur, as previously discussed, as well as an expansion event within Eurasia dated to 0.2535 Ma (Table 17.2). However, this within-Eurasia expansion event is not cross-validated by any other locus, so it will be ignored in this and all subsequent analyses. In addition to the two expansion events, there are 16 inferences of restricted gene flow involving Eurasian populations that have estimated ages older than 0.1774 (Table 17.2; Templeton 2002). The log-likelihood ratio test of the null hypothesis that all 18 of these times are no older than the most recent out-of-Africa expansion event yields a chi-square statistic of 131.56 with 18 degrees of freedom with a p value of less than $10^{-17}$. Excluding the inference of restricted gene flow associated with the outlier gene MX1, the chi-square is reduced to 118.18 with 17 degrees of freedom, which still yields a p value of less than $10^{-17}$. Hence, the genetic data overwhelmingly reject the out-of-Africa replacement hypothesis. There is no doubt that this spread of humans with many anatomically modern traits resulted in interbreeding with at least some Eurasian populations.

The NCPA results indicate that expansions coupled with interbreeding were not the only source of genetic contact between African and Eurasian populations during the Pleistocene. Figure 17.2 shows the gamma distributions for all the inferences of restricted gene flow among African and Eurasian populations from Table 17.2 and from the previous analyses (Templeton 2002, 2004a). Almost all of the older inferences of gene flow indicate that it was restricted by isolation by distance, but some of the more recent inferences (often not datable by phylogenetic techniques) indicate some long-distance dispersal as well. This means that this genetic interchange was mostly due to short distance movements to neighboring demes (other than the three major population-level expansions). Moreover, there were no major, persistent barriers (at least on a timescale of thousands to tens of thousands of years) between human demes in Africa and Eurasia, so copies of genes could spread throughout Africa and Eurasia via gene flow, using local demes as stepping-stones to cross vast distances over many generations.

A glance at Figure 17.2 versus Figure 17.1 reveals a dramatic difference; the inferences of range expansion are clustered, but the inferences of restricted...
gene flow form a continuum across the Pleistocene, with the exception of the inference associated with the outlier gene MX1. Because this is the only gene that indicates gene flow in the Pliocene, this inference is not cross-validated and therefore discarded. The continuum defined by the remaining genes is expected if gene flow restricted by isolation by distance were a recurrent evolutionary force throughout the Pleistocene, with no lengthy interruptions. Hence, the factors that allowed *H. erectus* to spread out of Africa nearly 2 Ma also allowed humans to go back and forth between Africa and Eurasia, at least in the stepping-stone sense.

Because gene flow is a recurrent evolutionary force, there is no expectation of different inferences of restricted gene flow from the various genes to be temporally concordant, in contrast to historical events such as rapid range expansions. However, information can be combined across multiple loci to define the probability of the time at which recurrent gene flow commenced among populations in a geographical area as (Templeton 2004a):

![Figure 17.2](image)

The distributions for the ages of the youngest clade contributing to a significant inference of restricted gene flow, primarily with isolation by distance. The $x$-axis gives the age in millions of years before present, and the $y$-axis gives the gamma probability distribution, $f(t)$. The genes or DNA regions yielding these distributions are, as ordered by their peak values of $f(t)$ going from left to right: Xq13.1, *MSN/ALAS2*, *HFE*, *FIX*, *HFE*, *G6PD*, *βHb ECP*, *RRM2P4*, *EDN*, *PDHA1*, *CYP1A2*, *FUT2*, *FUT6*, *FUT2*, *CYP1A2*, *CCR5*, and *MX1*. The curve for *MX1* is shown in a dashed line to emphasize its outlier status (Templeton 2002). Several other inferences of restricted gene flow that were too recent to date phylogenetically are not shown and not used in the analyses.
Setting Eq. (17.6) equal to 0.95 and solving for \( T \) defines the time at which we can be 95% confident that gene flow was established in a geographical area. Applying Eq. (17.6) to all the inferences of African–Eurasian gene flow found in Table 17.2 and in the previous analysis (Templeton 2002) yields a 95% confidence that gene flow was established between African and Eurasian population by 1.8694 Ma. Excluding the outlier \( MX1 \), the 95% confidence time is reduced to 1.4605 Ma. Thus, recurrent gene flow between African and Eurasian populations goes back to the lower Pleistocene and was established at or shortly after the initial spread of \( H. erectus \) out of Africa.

The inferences of ancient and recurrent gene flow punctuated by major population movements out-of-Africa coupled with interbreeding are consistent with the fossil record. Many fossil traits display a pattern of first appearing in Africa and then spreading throughout Eurasia (Stringer 2002), whereas other traits display a pattern of regional continuity (Wolpoff et al. 2000; Wu 2004). These two patterns are often regarded as mutually exclusive alternatives, but they are not under a model of genetically interconnected populations and no total replacement. As long as there is genetic interchange among populations, the Mendelian mechanisms of recombination and assortment allow different traits influenced by different genes to have different evolutionary fates. Some traits could have spread due to the joint actions of gene flow, admixture, and natural selection, whereas other traits may not have spread as rapidly or not at all due to a lack of selection or due to local selective pressures. Recurrent gene flow and admixture therefore provide the genetic interconnections that explains all of the fossil trait patterns during this time period, as well as current distributions of genetic variation in humans (Eller 1999; Eswaran 2002; Dugoujon et al. 2004). This model of gene flow and interbreeding also explains why current genetic variation in human populations does not fit an evolutionary tree model in which different human populations are treated as distinct “branches” on an evolutionary tree. Although the human evolutionary genetic literature is filled with portrayals of human populations as branches on a tree, none of these population evolutionary trees actually fit the genetic data when tested (Templeton 1998, 2003). Instead, patterns of genetic differentiation among current human populations fit an isolation-by-distance model much better than a tree model (Templeton 1998, 2003; Eller 1999, 2001).

Figure 17.3 summarizes the cross-validated, statistically significant conclusions from the NCPA based on 25 genes or DNA regions. The figure per se is
identical to Figure 1 in Templeton (2002) that summarized the cross-validated inferences from NCPA based on 10 genes or DNA regions because the expanded data set strongly reconfirmed the original inferences. The difference between the figures is in the text associated with the figure. In particular, the expanded analyses yields three new important inferences: (a) there was an expansion out-of-Africa about 1.9 Ma, (b) recurrent gene flow was established among Old World hominins by at least 1.46 Ma, and (c) the Acheulean expansion involved inter-breeding and not total replacement.
### 17.2.2 Ancient DNA analyses

The NCPA overwhelmingly indicates that the most recent out-of-Africa range expansion event was not a total replacement event. This does not mean, however, that there was interbreeding with all Eurasian populations. The possibility of some local replacement is compatible with the NCPA, but that possibility is not testable with NCPA because its inference is limited to historical populations that have left some genetic descendants in present-day populations. One way of addressing this possibility is to extract DNA from hominin fossils. The approach of using ancient DNA has been applied to the possibility of replacement of European Neanderthals by more anatomically modern forms between 30,000 and 40,000 years ago.

Working with ancient DNA is difficult. Because mtDNA is much more abundant than nuclear DNA, all DNA so far obtained from human fossils is mtDNA. Studies have revealed that Neanderthal mtDNA represents a unique mitochondrial lineage that is distinct from present-day human mtDNA and from the mtDNA found in fossil, anatomically modern specimens from comparable time periods. This pattern has been interpreted to mean that there was no or extremely little interbreeding between Neanderthals and their more anatomically modern contemporaries, and that Neanderthals perhaps represent a distinct species from *H. sapiens* (Krings et al. 2000; Caramelli et al. 2003; Knight 2003; Currat and Excoffier 2004; Serre et al. 2004). However, there are difficulties with these conclusions from ancient DNA studies.

First are technical difficulties. Ancient DNA is subject to damage over time, and resulting lesions can create artifactual substitutions (Caldararo and Gabow 2000; Hansen et al. 2001). One test for artifactual substitutions makes use of the considerable age range found in the Neanderthal fossils used as sources for DNA. If the apparent divergence is real, then the oldest Neanderthal samples should tend to be closest to current human mtDNA because they are temporally closer to the common ancestral sequence for Neanderthal and modern human mtDNA. In contrast, if DNA damage has made a large contribution to the apparent divergence, then the oldest Neanderthal sequences should be the farthest from that of modern humans. The later pattern is true (Gutierrez et al. 2002). Because the samples are small, one could argue that just by chance the oldest Neanderthal sequences just happened to come from an abnormally highly divergent lineage of Neanderthal mtDNA, but these results indicate that DNA damage cannot be discounted as a significant source of error in these studies. In addition, ancient DNA extracts induce artifactual mutations, both nucleotide substitutions and insertions/deletions, in a nonrandom fashion such that the same artifacts are independently created in controlled experiments (Pusch and Bachmann 2004).
Moreover, many of the sites at which these artifacts repeatedly occur are the same sites observed in Neanderthal mtDNA divergence (Pusch and Bachmann 2004). These results indicate that great caution should be exercised in interpreting ancient DNA sequence data.

Second, all the ancient DNA studies on human fossils have been confined to mtDNA, a molecule that has some unusual patterns of mutation and nucleotide substitution. Most of the analyses of Neanderthal mtDNA have ignored this fact. When the best fitting maximum-likelihood model of mtDNA evolution is used, there is no statistically significant support for a branch separating the Neanderthal sequences from modern human sequences even when all the sequences are assumed to be completely valid (Gutierrez et al. 2002).

Third, mtDNA is incapable biologically of completely reflecting a population’s evolutionary history and of rejecting the hypothesis of admixture. MtDNA is sensitive to only female-mediated gene flow and can totally miss even extensive interbreeding mediated through males. For example, Figure 17.3 shows a cross-validated expansion out-of-Asia that occurred after the most recent out-of-Africa event. This expansion event is marked by autosomal DNA and Y-DNA. This expansion event was not detected by mtDNA, even though the largest samples, the best-geographical coverage, and the greatest genetic resolution existed for mtDNA in this area. This implies that this expansion event was primarily mediated by males coming out of Asia and was invisible to mtDNA. A more recent example is provided by the Lemba of southern Africa, who show evidence for extensive interbreeding between Jewish and Bantu populations for Y-DNA and nuclear DNA but none for mtDNA, a pattern consistent with their oral history of a group of Jewish men coming down the east coast of Africa and then mating with local Bantu women to start the Lemba people (Soodyall 1993; Thomas et al. 2000; Wilson and Goldstein 2000). Thus, mtDNA alone is biologically incapable of detecting all admixture events, even those such as the Lemba in which the degree of admixture was 100%.

Fourth, as mentioned in the introduction, the evolutionary history of a single gene or DNA region should never be equated to the evolutionary history of a population. A glance at Figure 17.3 reveals that much of humanity’s recent evolutionary history is not detected at all by mtDNA. One needs multiple loci to obtain an accurate reconstruction of evolutionary history (Wall 2000) and to protect against false inferences due to evolutionary stochasticity and natural selection skewing the results of a particular gene (Templeton 2002, 2004b). In this regard, human mtDNA is known to have been subject to natural selection (Templeton 1996; Curnoe and Thorne 2003; Elson et al. 2004), and this provides alternative interpretations to the fossil mtDNA patterns. For example, suppose that there was much female-mediated gene flow between Neanderthals and
modern humans, but that in addition there was selection against the Neanderthal mtDNA haplotypes in a hybrid or predominantly modern human nuclear background, a phenomenon known as cytonuclear incompatibility. In such a situation, even high levels of interbreeding could go undetected by mtDNA.

Fifth, even if mtDNA is treated as neutral and gene flow was through Neanderthal females, the small sample sizes preclude the ability to dismiss significant amounts of gene flow between Neanderthals and moderns (Nordborg 1998; Wall 2000; Relethford 2001a; Currat and Excoffier 2004; Pearson 2004). Currat and Excoffier (2004) produce by far the smallest estimate of the amount of female-mediated gene flow compatible with the mtDNA assuming neutrality; an upper limit of 0.1%. However, even this is not an insignificant amount of gene flow. Within population genetics, gene flow is regarded as significant if the product of the population size times the migration rate is equal to or greater than one (Crow and Kimura 1970). Thus, if Neanderthals had a population size equal to or greater than 1,000, there could be biologically significant gene flow. Interestingly, the simulations used by Currat and Excoffier (2004) to obtain the limit of 0.1% started with a Neanderthal population size of 35,000. Hence, even this conservative evaluation indicates that biologically significant, female-mediated gene flow of neutral mtDNA cannot be rejected.

For the above reasons, the ancient mtDNA results should be regarded as suggestive of replacement of Neanderthals, but not conclusive. This inconclusive state will persist until ancient DNA technology can deal adequately with genetic artifacts and can be used to study multiple genes rather than just mtDNA. In the meantime, the fossil record (Duarte et al. 1999) and archeological record (Kaufman 2001) are perhaps more informative than the ancient DNA studies on the fate of the Neanderthals.

17.2.3 Demographic inferences from coalescent analyses

Because the coalescent process is influenced by basic demographic parameters, such as population size and population growth rates, coalescent analyses have been extensively applied to reconstruct past human demography. These analyses can both reinforce and complement the inferences made from NCPA. For example, NCPA infers a population range expansion out-of-Africa around 100,000 years ago, but it only identifies this as an expansion in geographical range, and not necessarily as an expansion in population size. One of the first molecular genetic studies that indicated that this time period was also marked by a population size expansion was the work of Rogers and Harpending (1992). They showed
that a sudden increase in population size induces a distinct peak (wave) in the
distribution of pairwise nucleotide differences (the mismatch distribution) under
a neutral model of evolution in a DNA region with little to no recombination
and with each new mutation occurring at a distinct nucleotide site (the infinite
sites model) and uniformly across sites. They found such a distinct peak in
human mtDNA, which apparently marks a rapid increase in human population
size. Computer simulations of the coalescent under an infinite sites model
revealed a good fit to the empirical curve with a population size expansion
between 60,000–120,000 years ago (later revised to 30,000–130,000 years ago by
(Harpending and Rogers 2000), which corresponds quite well to the most recent
out-of-Africa range expansion shown in Figure 17.3. They also showed that
Tajima’s D statistic (Tajima 1989a), an infinite sites model statistic that is sensitive
to past population bottlenecks, was consistent with their conclusions of growth
from a much smaller population size in the past.

Rogers (1992) recognized that the mismatch analysis and accompanying
computer simulations were based on a highly unrealistic model for mtDNA
evolution: the infinite sites model. As mentioned in the previous section,
mtDNA has some unusual patterns of DNA substitution that seriously violate
the infinite sites assumption. Rogers (1992) showed that the error introduced by
this assumption was only about 3% and perhaps less than 1%, a conclusion
reinforced by subsequent work (Rogers et al. 1996). In contrast, Yang (1997)
found that the estimates of ancestral population sizes are very sensitive to
mutational assumptions, and Schneider and Excoffier (1999) showed that one
needs to increase the mismatch expansion times by 10–20% larger than that
indicated under the infinite sites model. The reason for this discrepancy in how
much error is induced is that there are multiple alternatives once one leaves the
model of infinite sites and uniform mutation rates. Depending on the alternative
chosen, one can obtain trivial to substantial errors from infinite sites, uniform
mutation-based statistics. Unfortunately, our knowledge of the appropriate
model for DNA evolution is rather limited at present, and moreover, different
DNA regions normally deviate from the uniform infinite sites model in region-
specific ways (Fullerton et al. 2000; Templeton et al. 2000), so there is no single
alternative. We do know now that the deviations from the uniform infinite sites
model are much more extreme than those taken in account by Rogers (1992) and
Rogers et al. (1996). To illustrate how misleading the infinite sites model can be in
a case where the answer is known, Templeton et al. (2000) applied a standard,
infinite-sites estimator of recombination to human mtDNA. The infinite-sites
statistic detected 413 recombination events uniformly distributed over the
mtDNA genome. The problem is, mtDNA does not recombine! Thus, the con-
clusions based on infinite-site statistics can be egregiously wrong. This was also
noted in Section 17.2.2 concerning the evolutionary placement of Neanderthal mtDNA (Gutierrez et al. 2002). Such discrepancies can also be found in human nuclear DNA (Templeton et al. 2000). Unfortunately, infinite site statistics still dominate the human (and nonhuman) literature. Until we have a greater knowledge of appropriate DNA models and the development of more statistics and simulation packages that do not assume an infinite sites model and uniform mutation rates (and the willingness of investigators to thoroughly examine their data in order to choose an appropriate mutational model for their specific DNA region), demographic conclusions should be regarded as subject to an unknown, but potentially substantial, degree of error.

The statistics used for demographic inference are also sensitive to natural selection (Tajima 1989a, b; Harpending and Rogers 2000). A dramatic indicator of the potential of selection to bias results can be seen in a coalescent analysis to estimate the amount of population size expansion in recent human evolution based on 612 single-nucleotide polymorphisms (SNPs) (Wooding and Rogers 2002). The SNPs were subdivided into three categories: coding nonsynonymous, coding synonymous, and noncoding DNA regions. The coding nonsynonymous SNPs are generally regarded as the ones most likely to be subject to natural selection. Wooding and Rogers (2002) obtained maximum-likelihood estimates of the ratio of the human population size in the most recent epoch to the population size in an earlier epoch for all three SNP categories. The coding nonsynonymous SNPs yield a ratio of 9,900, implying massive population growth. In contrast, the ratios were 0.4 for coding synonymous and 0.6 for noncoding SNPs, implying little or no change in population size. Thus, dramatically different inferences arose among these categories of SNPs that should be differentially sensitive to natural selection. Note also that the inferences least likely to be affected by selection indicated little or no change in population size.

Because selection and demography are confounded in the coalescent-based statistics used for demographic inference, it is important to use cross-validation across multiple loci, just as was done in NCPA. However, cross-validation is probably less effective in protecting against false demographic inferences than in protecting against false phylogeographic inferences using NCPA. At the molecular level, the two most common patterns of selection that are reported in the literature are negative (purifying) and positive (directional) selection (Fay et al. 2001; Miller et al. 2004). Negative selection occurs when mutations at a locus or DNA region are primarily a mixture of neutral and deleterious mutations such that the role of natural selection is to eliminate the deleterious mutations, thereby preserving the important structural features controlled by the gene. About 70% of all nucleotide substitutions leading to amino acid changes are deleterious in humans and chimpanzees (Enard and Pääbo 2004). Because
deleterious mutations are eliminated by natural selection, they play little if any role in shaping the haplotype tree (Barton and Etheridge 2004). The NCPA inferences arise from the haplotype tree, and therefore purifying selection does not confound NCPA inferences. The same is not true for population size inferences. The selective elimination of deleterious mutations in a DNA region with little or no recombination [such as mtDNA, and now, much of the human nuclear genome as well (Reich et al. 2002)] mimics the effects of a simple reduction in population size in the past (Charlesworth et al. 1995). Unfortunately, the other most common form of selection, directional or positive selection, also mimics a past reduction in population size (Tajima 1989a). Directional selection causes a selective sweep, eliminating all previous variation within the DNA region that does not recombine with the selected mutation. Since NCPA makes inferences only on the variable part of the haplotype tree with no inferences possible past the coalescent time to a common ancestral molecule, the actual inferences made by NCPA are not affected by selective sweeps except in the rare case of actually sampling a locus in the middle of such a sweep. Thus, while NCPA is expected to be robust to both negative and positive selection, population size inferences are affected by both and in the same direction. This shared bias from the two most common forms of reported selection diminishes the effectiveness of cross-validation for demographic inferences. Recall from above that deviations from the infinite site model also result in underestimates of time and population size (Schneider and Excoffier 1999). Hence, many estimates of past population sizes and their ages are probably underestimates, even when based on multiple loci.

Two other forms of selection can cause a demographic bias in the opposite direction. The first is balancing selection in which natural selection maintains two or more haplotype lineages or alleles in a polymorphic state. Balancing selection can extend coalescence times into the past, which inflates apparent population size. The coalescent processes within each selected haplotype lineage obey neutral coalescent theory (Hudson 1990), so balancing selection has only a limited impact on the haplotype tree (Barton and Etheridge 2004). Because of the nested clade design used in NCPA, most clades would be unaffected by balancing selection, with only those clades containing different selective lineages (typically only those at the highest level of nesting) being influenced. Hence, NCPA inferences are mostly robust to balancing selection, but with some exceptions which should be dealt with via cross-validation. The final type of selection is diversifying selection due to local adaptations in a spatially heterogeneous environment. Such selection also maintains polymorphism in the species as a whole, and hence mimics the biases seen with balancing selection. However, unlike balancing selection, this type of selection can create strong geographical correlations
and thereby lead to false inferences for NCPA. Multilocus cross-validation is therefore important in both demographic inference and NCPA to protect against the effects of balancing and spatially diversifying selection.

Cross-validation was used for the mtDNA inference of a population size expansion between 30,000 and 130,000 years ago. Although some nuclear loci do indeed cross-validate this result (Marth et al. 2003; Marth et al. 2004), others do not (Harpending and Rogers 2000). Recall also that the coalescent analyses of the SNPs least likely to be subject to selection did not indicate a recent population size expansion (Wooding and Rogers 2002). Similarly, an analysis of 10 non-coding DNA regions (to minimize selection) found no evidence for significant population size expansion in Africa and only nominal significance (without correcting for multiple testing) in Eurasian samples (Pluzhnikov et al. 2002). Harpending and Rogers (2000) attempt to salvage the idea of a recent population size expansion by arguing that balancing selection is pervasive in the human nuclear genome, which, as discussed above, inflates the apparent population sizes of the past and thereby reduces the apparent amount of population growth. This assumption is inconsistent with recent surveys of natural selection that found evidence for positive or balancing selection in only 8 of 132 genes in Eurasians and 0 out of 132 in Africans (Akey et al. 2004). This explanation is also inconsistent with studies indicating little population size expansion that use noncoding or silent variation that is unlikely to be subject to balancing selection (Pluzhnikov et al. 2002; Wooding and Rogers 2002). Thus, the evidence for a small ancestral human population size is mixed and a coherent picture has yet to emerge (Pluzhnikov et al. 2002; Wooding and Rogers 2002).

A separate demographic issue concerning the most recent out-of-Africa range expansion shown in Figure 17.3 was whether it involved a large population coming out of Africa or a small one. Many genetic surveys show that haplotype diversity is generally much lower in Eurasian populations than in African, and that Eurasian populations show more linkage disequilibrium (non-random associations between polymorphic sites that can be induced by both small population size and admixture) than African populations. These patterns indicate that the African population coming into Eurasia was relatively small initially (Tishkoff and Verrelli 2003), yet dominating in the subsequent interbreeding with Eurasian populations. These patterns could also mean that the Eurasian population was smaller than the African population during most of the Pleistocene (Relethford 1998) and would be compatible with a larger Eurasian input. Regardless, these genetic observations raise another serious bias to estimating ancestral human population sizes prior to the most recent out-of-Africa expansion event shown in Figure 17.3; namely, the strong bias in many human genetic surveys toward non-Africans (Tishkoff and Verrelli 2003).
By being biased for Eurasians, the small ancestral sizes reported by many may reflect the size of the population coming out of Africa or smaller Eurasian populations throughout much of the Pleistocene rather than the global human population size at that time. Indeed, studies having a large sample of Africans tend to have much larger estimates of ancestral human population size (Tishkoff and Verrelli 2003), even to the extent of showing no significant or moderate population size growth from the past to the present within Africa (Pluzhnikov et al. 2002; Adams and Hudson 2004).

A major limitation of many of the coalescent estimators relative to NCPA is that they are based on simulating various demographic models and measuring the fit to the data. Thus, the inference universe is limited to the scenarios that were simulated. Even an excellent fit to the data does not insure that the simulated model is the right one because often several models can fit the data well. For example, Adams and Hudson (2004) pointed out that one of the African populations they studied fit both models of constant population size and a variety of growth models, including fivefold growth beginning no earlier than 36,000 years ago. The real weakness of the simulation approach is that it is never possible to simulate all possible scenarios; so, many alternative hypotheses are never evaluated at all. For example, none of the simulations performed in the papers referred to in this section take into account the fact that the expanding populations out of Africa interbred with Eurasian populations (recall that the hypothesis of total replacement was rejected with a $p$ value less than $10^{-17}$), nor the long-continued history of gene flow constrained by isolation by distance between African and Eurasian populations that extends to 1.46 Ma with 95% confidence. Thus, the demographic scenarios with the most compelling genetic evidence have never even been considered in these simulations.

Based on the literature from the early 1990s (reviewed in [Hawks et al. 2000]), it is commonplace to find many papers citing a figure of 10,000 for the ancestral human population size and even regarding this population size as a well-established constraint in interpreting human evolution (Pearson 2004). Given that inappropriate DNA models, positive and negative selection, and sampling biases, all conspire to underestimate this ancestral human population size; it is now obvious that 10,000 is an underestimate. Indeed, the more recent studies that most thoroughly try to eliminate these biases as mentioned above usually detect no significant-to-moderate population growth from the recent past to the present, particularly in Africa, implying that the human ancestral population size was much greater than 10,000.

The error of regarding the figure of 10,000 as a given constraint on recent human evolution is sometimes amplified by regarding this figure as an estimate of the census size of the human population at that time (Pearson 2004).
No coalescent estimator of population size estimates census size; rather, they all estimate the inbreeding effective size. A coalescent event establishes a relationship known in population genetics as identity by descent. Population genetics also makes frequent use of an idealized population characterized by self-compatible hermaphrodites that are randomly mating in an unsubdivided population, all with the same average number and variance of offspring (as described by a Poisson distribution), no selection, discrete generations, and constant population size. These assumptions allow one to derive formulae that describe how a variety of population genetic parameters evolve, including the probability of identity by descent. Of course, real populations, and humans in particular, deviate from this idealized population. Population genetics uses the concept of an effective population size to create a common reference for all real populations. One chooses a particular genetic parameter, such as the probability of identity by descent or the variance in allele frequency from one generation to the next and observes how this parameter evolves in the real population. Then, one calculates the size of an idealized population that would yield the same parameter values as observed in the real population. The calculated size of the idealized population that mimics the evolution of the parameter of interest in the real population is known as an effective population size. It has long been known that deviations from the idealized population can create quantitatively and qualitatively different deviations from census size for different population genetic parameters. For example, an expanding population size tends to decrease the inbreeding effective size (the effective size for the genetic parameter of probability of identity by descent) but increase the variance effective size (the effective size for the genetic parameter of the variance in allele frequency from one generation to the next) (Crow and Kimura 1970). It is incorrect to define “the” effective size as the number of breeding individuals, as done for example by Pearson (2004). Rather, there are many different effective sizes, depending on the genetic parameter of interest. The effective size estimated from coalescent-based approaches is the inbreeding effective size.

There is no expectation for the inbreeding effective size to equal the census size, nor to equal other effective sizes. The absurdity of equating an inbreeding effective size to a census size is shown by the fact that the inbreeding effective size of Africans alone is larger than the inbreeding effective size of Africans and Eurasians together, as pointed out above. Hence, if we equated these effective sizes to census sizes, we would conclude that more people lived in Africa alone than in Africa plus Eurasia combined! Under realistic circumstances, these effective sizes can differ from one another and from census sizes by orders of magnitude (Templeton and Read 1994). For example, models of demes interconnected by gene flow but with some local extinction and recolonization
are applicable to many species, including humans. Such “metapopulation” models can greatly reduce inbreeding effective size relative to census size (Wakeley 2004). Indeed, Eller (2002) showed that such a demographic model allows the human census size to be approximately 300,000 through the Pleistocene and still be compatible with an inbreeding effective size of 10,000. Consequently, arguments that a population size of 10,000 is too small to support an African and Eurasian distribution (Pearson 2004) are without substance even if one accepts an inbreeding effective size of humans of 10,000—a conclusion itself that is highly questionable. Thus, no coherent picture has emerged from genetic studies concerning the size of Pleistocene hominin populations, but a nongenetic estimate of around half a million individuals (Weiss 1984) is consistent with many of the latest genetic results and the calculations of Eller (2002).

17.3 Natural selection and adaptive evolution in hominin populations

17.3.1 Adaptive evolution at the protein level

Coalescence theory and phylogenetics provide a variety of tests to infer the presence and type of natural selection operating at the molecular level, particularly in protein-coding genes (Bamshad and Wooding 2003; Tishkoff and Verrelli 2003; Enard and Pääbo 2004). Indeed, the great abundance of sequence data, not only in humans but in closely related species (outgroup data are frequently required for many of these tests of selection), have allowed massive screening throughout the human genome to identify those genes and DNA regions that were specifically subjected to positive, directional selection in the lineage leading to modern humans and thus were involved in the adaptive transformation of the human species.

Clark et al. (2003) found 178 genes out of 7,645 protein coding genes that had significant ($p < 0.01$) evidence of positive selection with standardized ratios of nonsynonymous to synonymous nucleotide substitutions with a model that allowed the amino acid sites to either be neutral, under negative selection, or under positive selection specifically on the branch of the gene tree leading from the common ancestor of humans and chimpanzees to present-day humans. Many of the genes that were under positive selection in the human lineage involved sensory perception, particularly olfaction and hearing. Several other genes under positive selection were involved in amino acid catabolism, and may have been selected as the dietary habits of the human lineage diverged from the human–chimp ancestral state. Many developmental genes were also under positive
selection in the human lineage, and these fell into two main categories: skeletal development and neurogenesis. Looking specifically at genes involved in the nervous system, Dorus et al. (2004) found accelerated rates of amino acid substitutions in the primate lineages leading to humans, with the highest rate being found in the branch leading from the common ancestor of humans and chimpanzees to humans. Vallender and Lahn (2004) reviewed the literature on those genes showing positive selection in the lineages leading to humans, and found that the vast majority of them fell into just a few functional categories: host–pathogen interactions, reproduction, dietary enzymes, sensory perception, central nervous system (CNS) functioning, and brain anatomy.

The above studies indicate that positive selection played an important role in the adaptive evolution of the human lineage, but these studies only localize that selection into the time period between the common ancestor of humans and chimpanzees to the present. One way of localizing the positive selection to the Pleistocene is to look for evidence of a selective sweep. As mentioned in Section 17.2.4, positive selection induces a selective sweep of the genetic variation linked to the favored mutation. This in turn leaves a signature in the haplotype tree that is gradually lost over time, limiting this approach to only the latest stages of human evolution (Przeworski 2002). Such selective sweeps have been found for the monoamine oxidase A gene (Gilad et al. 2002), involved in the functioning of the CNS, and in FOXP2, a gene coding for a transcription factor involved in speech and language development (Zhang et al. 2002). An alternative approach is to look for evidence of positive selection confined to Eurasian populations. This would imply not only some sort of local selection involved with the colonization of Eurasia but also confine the time period to the last 2 Myr. Akey et al. (2004) surveyed 132 genes and found evidence for positive or balancing selection in eight of them in Eurasian populations, but none in African populations. Similarly, a survey of 624 autosomal loci found evidence of positive selection in 13 genes, and of these 12 exhibited only a Eurasian signature of positive selection (Storz et al. 2004). These studies indicate that human populations had to adapt to novel selective features associated with the Eurasian continent. Thus, the expansions out-of-Africa shown in Figure 17.3 were also associated with altered selective regimes that influenced human adaptive evolution.

17.3.2 Adaptive evolution of hominin facial morphology

When selection operates upon a morphological trait, it generally alters other traits as well due to shared underlying developmental process and genes. These
correlated responses to selection can be captured by a quantitative genetic variance/covariance matrix (Ackermann and Cheverud 2004). If a set of morphological traits is evolving neutrally or are neutral correlates of another, unmeasured trait under selection, then their coordinated change should reflect the variance/covariance matrix (Lande 1980). Deviations from this expected correlational structure indicate the action of natural selection. Ackermann and Cheverud (2004) estimated the variance/covariance matrices for several facial features from living human, chimpanzee, and gorilla populations, and used all three of these living models to detect selection on facial features in hominin fossils. The results within the *Homo* fossils were robust to all three models, showing that most facial evolution was neutral between 1 and 2 Ma. In contrast, selection was necessary to produce the *Homo* face from that of a gracile autralopith, with selection being exerted to increase the relative portions in the upper face and orbit, weak to no selection to increase the midface/nasal region, and selection to reduce the lower orbits and zygomatics. Hence, in terms of the human face, the most important adaptive evolution occurred leading up to *Homo*, with little evidence for selection during the Pleistocene. This is consistent with the idea that the development of cultural inheritance could have released many morphological traits of humans from the effects of selection (Lynch 1990; Ackermann and Cheverud 2004). If this explanation is true, it would imply that hominins increasingly relied on technology near the time of the spread of *H. erectus* out of Africa.

### 17.4 Conclusions

The first major event in hominin evolution over the past 2 Myr was a range expansion out of the African homeland into Eurasia. This expansion event may have been triggered by an increase in the importance of cultural inheritance as a means of adapting to the environment. Regardless, the time of this expansion marked an era of relaxed selection on some morphological features. Although some morphological selection may have been relaxed as hominins expanded into Eurasia, natural selection nevertheless operated strongly and swiftly upon many genes as hominins adapted to novel environmental circumstances in Eurasia.

Once humans acquired the ability to disperse out-of-Africa, they obviously retained the ability to disperse back in, as recurrent gene flow, constrained by isolation by distance, was established by at least 1.46 Ma. Because this initial colonization of Eurasia left detectable genetic signatures in present-day Eurasian
populations, there has been a continuous Eurasian hominin presence for at least the last 1.9 Myr. Africa still remained the center of hominin evolution, however, with the next major Pleistocene event being the development of the Acheulean culture and its subsequent spread out of Africa. This spread was both cultural and demic, with the expanding peoples from Africa interbreeding with at least some of the Eurasian populations with whom they came into contact.

Recurrent gene flow continued after the Acheulean expansion, still with isolation by distance. Such a population structure would allow some traits to spread throughout all of humanity while other traits were locally restricted. This pattern of local restriction is observed in multiple selective sweeps and balancing selection confined to Eurasia despite recurrent gene flow. No coherent picture has emerged from genetic studies concerning the size of the middle Pleistocene hominin population, but a nongenetic estimate of half a million individuals (Weiss 1984) is consistent with many of the latest genetic results.

Many anatomically modern traits first evolved in Africa, and then peoples with these traits spread out of Africa about 100,000 years ago. As with the Acheulean expansion, this expansion involved interbreeding with at least some Eurasian populations and was overlaid upon a recurrent pattern of genetic interchange. Modern traits could spread by this demic expansion coupled with interbreeding as well as through gene flow with isolation by distance. As before, this population structure allows locally adaptive traits to maintain their regional continuity. However, it is possible that some Eurasian populations were replaced. The potential replacement of the Neanderthals by people with more modern traits cannot be inferred conclusively from current genetic evidence.

After this last major out-of-Africa expansion event, humans expanded into previously unoccupied parts of the world: northern Eurasia, Australia and the Pacific, and the Americas (Figure 17.3). There was also a major expansion of humans out of Asia back toward Europe and Africa that was primarily male mediated (Figure 17.3). Wherever humans spread, genetic interconnections were maintained with other human populations, at least on a timescale of several thousands of years. These connections of gene flow were restricted through isolation by distance, although long-distance dispersal becomes more important in very recent times (Table 17.2). Even today geographical distance remains the primary indicator of the degree of genetic differentiation between local human populations (Eller 1999). The relationships among current human populations therefore cannot be represented as a tree of populations but rather an intertwined trellis (Figure 17.3). As a result, humanity evolved into its modern form as a single evolutionary lineage but with some geographical differentiation at any given time.
Acknowledgments

I am grateful for support by NIH grant GM065509. I also want to thank Rebecca Ackermann and James Cheverud for kindly making their manuscript on hominin evolution available to me prior to its publication.

References


Population biology and population genetics of Pleistocene hominins

Templeton AR, Boerwinkle E, Sing CF (1987) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction
Abstract

The idea that evolution takes place at varying rates has a long history, as does the argument that evolutionary change is concentrated at points of speciation. It was however not until 1972, with the publication of a seminal paper by Eldridge and Gould, that these two ideas were brought together and became widely discussed under the term punctuated equilibrium and viewed as an alternative to “phyletic gradualism.” If speciation is indeed the engine of evolution, it is necessary to ask what precisely species are. There is widespread acknowledgment that the Evolutionary Species Concept, first proposed by Simpson in 1961, reflects the reality of species: a species is a unitary evolving lineage. The debate about the definition of species is essentially concerned with how best to operationalize this concept. I show here that the closest operational definition is the so-called Phylogenetic Species Concept in which a species is the minimal cluster of individuals that is diagnosably distinct from other such clusters. There is, importantly, no such thing as an “amount of difference” necessary for species status. Speciation is usually assumed to be allopatric, but interesting cases have been made that other modes, such as sympatric, are not only feasible but also potentially important; speciation also occurs by hybridization. Nor can phyletic speciation be discounted. Speciation may occur very rapidly indeed. I finally discuss how different modes of speciation could be detected in the fossil record.

18.1 The PE revolution

It is probably true to say that until 1972 most paleontologists had interpreted their fossils as showing gradual change with time, and if they did not show gradual change it could be inferred: failure to demonstrate much or any change could be put down to the incompleteness of the fossil record. The year 1972 saw the appearance of the paper by Eldredge and Gould introducing the concept of
Punctuated Equilibria (hereinafter PE) in which they proposed that the major process of evolution is not, after all, one of gradual modification but a series of stops and starts: there are long periods when nothing happens, succeeded by short bursts of rapid change. The bursts of change, moreover, are not in the central, paleontologically most visible, population of the species but in small peripheral populations which have the capacity to invade the center and usurp the parent population.

Long ago, Simpson (1944) had recognized that evolutionary change occurs at different rates; along his rate spectrum he distinguished horotely ("normal" or average rates of evolution), bradytely (unusually slow), and tachytely (unusually fast). Early critics of PE assumed that it was merely a resurrection of these old Simpsonian categories. But it is not mainly about this: it is about stasis and speciation. Simpson was of course well aware of stasis; he described how in many cases a fossil species did not change at all over a considerable segment of time but remained the same. He was, too, aware of the importance of speciation. From the process of evolution along a single lineage (anagenesis), which he did seem to consider primary, Simpson (1944) carefully distinguished cladogenesis (evolution by splitting of one lineage into two). This typically results in two species where formerly there was one, so cladogenesis can be thought of as, broadly speaking, equivalent to speciation. What Eldredge and Gould had observed was that, in many cases, long-term stasis is followed by the sudden appearance of a different species which replaced the earlier one. The real innovation of PE was to insist that it is stasis, not anagenesis, that is the usual state of affairs, and that speciation, not anagenesis, is the engine of evolution.

Eldridge and Gould were perfectly aware, of course, that they had intellectual precursors. Three of these were especially influential. One of these was Richard Goldschmidt, a geneticist who in 1940 proposed the "hopeful monster" hypothesis, under which major evolutionary changes occur abruptly. The "hopeful monster" was not part of the original package, but it was espoused in later versions by Gould (1980, 2002). A mechanism for "hopeful monsters" was later proposed by Schwartz (1999).

Another inspiration for PE was Mayr who, in 1963, urged the Founder Principle as a mechanism of speciation. In Founder Principle theory, a small segment of a species becomes isolated from the rest (as for example by geographic or environmental change) and undergoes rapid genetic differentiation. He preferred this mechanism to the generally assumed view that a species, when divided, splits into two fairly equal halves which then diverge gradually. Schopf (1972), in his editorial introduction to the paper, traced the idea of Founder Principle back to Haldane in the 1930s and even as far back as Bernard in the 1890s, but there is no doubt that it was Mayr whose careful and detailed formulation of the principle
established it as a significant mechanism—or even the significant mechanism—of speciation.

The third major influence on the genesis of PE was another geneticist, Sewall Wright, who in a series of papers (see especially Wright 1968) formulated the principles of genetic drift, which is random change in gene frequencies. As he showed mathematically, later elaborated in detail by Kimura (1983), stochastic processes have a small but finite probability of resulting in the fixation of new genetic variants, even in the face of weak negative selection.

Where does this leave anagenesis? Long-term trends in evolution of course exist, but the problem is how to explain them. Osborne in 1936 ascribed evolutionary trends to orthogenesis, a sort of (internally controlled) urge to progress onward and upward, always in the same direction. Simpson (1944) effectively demolished orthogenesis, and instead proposed what he called orthoselection, in which natural selection forces fairly consistent change in one direction as long as local environments favor it. Eldredge and Gould (1972) suggested how PE might explain evolutionary trends, but the problem was more fully considered by Vrba (1980) whose “effect hypothesis” has become part of the theoretical core of PE. We will return to the matter of long-term trends in a later section.

So we have two apparently opposed models of the evolutionary process. Gradualism allows for cladogenesis to be a major evolutionary mode, though it does not mandate it; if, for example, lineage splitting does occur, it is under the gradualistic model followed by the slow mutual divergence of the two daughter lineages. Under PE, it is cladogenesis which is definitely the major evolutionary mode; when lineage splitting occurs, it is asymmetrical, and by the founder effect one daughter lineage diverges but the other generally does not. The major problem that lies at the heart of the difference is the nature of speciation.

### 18.2 The nature of species

#### 18.2.1 Species: theoretical concepts

Before we can even consider what speciation involves, we must very carefully delineate the nature of species. If we are to contemplate whether species, which are already regarded as the units of ecology, biogeography, and nowadays of conservation strategy, are also the units of evolution itself, then we must be clear exactly what species are. The importance of what we mean by the word “species” goes way beyond systematics, the field of study for which the term was invented.
It is unfortunate in one way, appropriate in another, that there is quite a multitude of different conceptions of what a species actually is. How many “species concepts” one distinguishes is a matter of how finely one is prepared to seek out the fine distinctions between different authors’ manner of phrasing. Thus Harrison (1998) distinguishes only 7 but Mayden (1997), and following him Hey (2001), as many as 24. Broadly, we may divide them into those which emphasize the absence of interbreeding and those which do not: “theoretical” and “operational” concepts in the terminology of Groves (2001). The division also corresponds, in broad terms, to “pattern” versus “process”: the pattern that we can observe versus the process which maintains it or brought it about in the first place.

For Dobzhansky (1937) and Mayr (1963), the nature of species lies in its reproductive isolation from other species. This view of reproductive isolation as the defining feature of species is called the Biological Species Concept (BSC). Species were defined by Mayr (1940) as “groups of actually or potentially inter-breeding natural populations which are reproductively isolated from other such groups,” while Dobzhansky (1950) summarized the situation by saying that the species is “the largest and most inclusive ... reproductive community of sexual cross-fertilizing individuals which share in a common gene pool.”

Reproductive isolation in turn depends on reproductive isolating mechanisms, which may be either premating or postmating. Premating mechanisms are those that prevent potential mates from meeting (seasonal or ecological), or from mating if they do meet (ethological), or from permitting sperm transfer if they actually get as far as mating (mechanical). Postmating mechanisms are what prevent gametes or zygotes from surviving if mating has nonetheless successfully occurred, or hybrids from surviving, or hybrids from breeding if they do survive.

Paterson (1978, 1980, 1982) argued that postmating isolation cannot be more than a symbol of species status: two sexually reproducing genotypes, A and B, do not avoid interbreeding with each other as such, instead A selects other A individuals and B selects other B. They do this by means of specific mate recognition systems (SMRSs): one individual emits a signal, to which another responds. A species, for Paterson (1978, 1982), is a population (or group of populations) whose members share a common SMRS: he calls this the Recognition Species Concept. Examples of SMRSs could be vocalizations, such as distinguish species of bush babies or galagos (Primates: Galagidae), whether sympatric species pairs (Nash et al. 1989; Masters 1991), or allopatric populations within what had previously been deemed unitary species (Bearder et al. 1995); or facial markings, together with the facial expressions, head flagging, and whole-body movements, that serve to emphasize these markings, as between species of the
genus *Cercopithecus* (Kingdon 1980, 1988). In principle, all one would have to do is to look at facial patterns in two given populations of animals and, if they are different, separate them as different species. But how much difference is “different”? And what if members of the two populations do, nonetheless, interbreed? The same question could be asked about vocalizations: how “different” do they have to be to qualify?

Templeton’s (1989) Cohesion Species Concept is another attempt to modify the BSC. In this concept, a species is defined by the genetic and demographic mechanisms which give it its cohesion: the genetic mechanisms are those which limit its gene flow with other species, and the demographic ones are those which constrain it to its “fundamental niche.” This tries to fuse the original BSC with the Recognition Concept while incorporating a general appreciation that a species must have its own ecological niche; but one would be hard put to use such a concept in practice.

The BSC and the attempts to modify it work very well in cases of sympatry. If two taxa are sympatric, then they are reproductively isolated: they are distinct species and no further discussion is possible. If they are parapatric, then equally they have the chance either to interbreed or not. But what if they are allopatric? Several authors have emphasized that the BSC is simply not applicable in cases of allopatry. Berlocher (1998) noted that the first author to point this out was A.R. Wallace himself in a paper written in 1864.

Mayr (1963; Mayr 1969) offered three criteria to decide whether two allopatric taxa are distinct species or merely subspecies. The amount of difference between allopatric taxa should be compared to that between

1. Sympatric taxa in the same group; or
2. The most divergent intergrading subspecies; or
3. Freely hybridizing populations in the same group

The first criterion founders on such things as the existence of sibling species, i.e., those that are nearly indistinguishable, which nonetheless may vary geographically. Both Dobzhansky and Mayr emphasized the importance of sibling species, and described cases in detail, but did not comment on how their very existence renders criterion (1) impossible. As for the second criterion, “degree of difference” is not strongly correlated with the propensity to interbreed: the two most distinctive species of baboons, the Hamadryas Baboon *Papio hamadryas* and the Olive Baboon *Papio anubis*, interbreed along their common geographic border, whereas the much less distinctive Chacma (*Papio ursinus*) and Yellow (*Papio cynocephalus*) Baboons are not known to interbreed along theirs. The third criterion depends on the ability to distinguish between primary and secondary
intergradation which, if the hybridization is too “free,” may be a very subjective matter indeed. When all is said and done, “amount of difference” is a very slippery concept, whether we consider morphology, vocalizations, DNA, or any other type of differences.

If we require that a species must be monophyletic (but see below), then we have another problem with the BSC and its modifications. The relationship between common descent and interbreeding has been explored by de Queiroz and Donoghue (1988), who conclude that they may or may not coincide. If they do not, then a species will not be monophyletic.

A different problem is a purely practical one: how do we know, in any given case, whether reproductive isolation really exists, or when we are dealing with SMRSs? Detailed field observations may inform us in the end, but these are not available in most cases—whether to test for the reality of SMRSs or for the existence of reproductive isolation.

Yet we have to agree that reproductive isolation, in the cases where it can be demonstrated, is absolutely decisive. Many authors, including Groves (1989), have been so impressed by this that they have treated “speciation” as equivalent to “the evolution of reproductive isolation.” And how this occurs is indeed an important question: Is reproductive isolation a by-product of divergence, as argued by Harrison (1998), or can it be its cause, as argued for example by the advocates of sympatric speciation?

18.2.2 Species: operational concepts

If one is working with preserved animals in a museum, or with the fossil record, the species criterion of the BSC—however much one tinkers with it—can be at best an inference, at worst misleading. As Nixon and Wheeler (1990) put it, what can be observed under such conditions can only be pattern, not process. This was observed more than 30 years ago by Sokal and Crovello (1970) and was the reason why they rejected the BSC, though they did not explicitly propose an alternative; it was left to Cracraft (1983, 1989, 1997) to propose the Phylogenetic Species Concept (PSC), under which a species is regarded as the minimal cluster of individuals that is diagnosably distinct from other such clusters.

The concept was called “phylogenetic” because, under it, species are the terminal points on a cladogram; hence they are the least inclusive phylogenetic units. But it stands to be misunderstood; Avise and Ball (1990), Mallet (1995), and Harrison (1998) all misread it as specifying that a species must be defined by apomorphies. This is not the case: a cladogram depicts sister-group relationships,
not ancestors, and a terminal need have no autapomorphic states. It is therefore perfectly possible under the PSC to have a species that is defined only by primitive retentions, a “living fossil” in the broad sense. It follows from this that a species may not be monophyletic: indeed, the implication of speciation by founder effect would be that, initially at least, up to half of all species are not. Baum (1992) refers to such nonmonophyletic ancestral forms as metaspecies.

There is also some disagreement over exactly what is distinct under the PSC. “Diagnosably distinct” (Cracraft 1983) may mean one diagnostic character state (“at least one”: Baum 1992) or it may require “a unique combination” (Nixon and Wheeler 1990; Christofferson 1995). In the main the case may be academic, but in the extreme case a single fixed heritable difference from its relatives still allows a population to be diagnosable, hence to be a species. But remember that fixed heritable differences may in fact be expressed only in one sex or at one stage in the life cycle. Under the PSC, therefore, species are units of genetic cohesion, not necessarily of reproductive cohesion, contra Kimbel and Martin (1993), because different phylogenetic species may still interbreed.

The PSC is eminently operational. While we can rarely or never be entirely certain that a given character state is fixed (i.e., at 100%), the PSC relies on the evidence to hand, while the BSC and its modifications demand that we go in for a great deal of inference. Paleontologists have little option but to use the PSC (contra Gould 2002 pp 785–789), and in reality neontologists usually have no option either.

18.2.3 What are different species concepts trying to say?

The Evolutionary Species Concept (ESC) was proposed by Simpson (1961): “A lineage . . . evolving separately from others and with its own evolutionary role and tendencies.” Subsequent authors have modified this, or drawn it out, in different ways; thus Christofferson (1995) explains that, as far as sexually reproducing organisms are concerned, a species is “a single lineage . . . genetically integrated by historically contingent events of interbreeding.”

The supreme importance of this concept is acknowledged by all commentators; it is the very essence of the species—why the species category is so vital (Mayden 1997) and why Ghiselin (1974) reified species as “individuals.” It is therefore a different level of concept from either the BSC or the PSC. The fact that Simpson’s definition (or Ghiselin’s gloss on it, for that matter) cannot be used in decision-making takes us back to the PSC as the means whereby we can recognize species.
18.2.4 Does species status necessitate a certain degree or quality of difference?

The literature on species, including in mammals, is full of statements implying that differences between species are of a different order or degree to those between subspecies, populations, or individuals within a species. Often, there are no explicit statements to this effect; rather, it is assumed that in some character or other, species differ grossly or qualitatively (however that is interpreted), whereas mere subspecies differ quantitatively: thus between and within species differences are not of the same kind.

Bohlken (1958), for example, used an allometric method. He found that on his bivariate plots of skull and horn measurements different species of Bovini generally have different slopes; and when by contrast he found that two of them (the tamaraw, Bubalus mindorensis, and the Indian wild buffalo, B. arnee) did not, he used this as a reason for combining them into one species. Imaizumi (1970) also used a variant of this allometric method with respect to Japanese and Chinese deer (Cervus nippon group); in this case, he plotted skull variables against an index of temperature on the hypothesis that different species would show different responses to Bergmann’s Rule whereas different subspecies within a species would respond similarly.

The idea that species ought to be more differentiated genetically than infraspecific groups has a long history. In a classic paper, Ayala (1975) compared values for Identity (Nei’s I) and Distance (Nei’s D) at various loci for various animal groups, finding that well-differentiated species pairs characteristically have higher values of D than do sibling and sister species which are higher than subspecies which in turn are higher than local populations. Thorpe (1983) collected an even larger amount of data, also using Nei’s D, and he too concluded that there is an average difference between species and infraspecific levels. Bradley and Baker (2001) tested genetic distances at different taxonomic levels using cytochrome b sequence data for a number of bat and rodent genera, and once again found that genetic distances do tend to increase in the expected sequence from interpopulational to interspecific; moreover, distances between presumed sister species were less than between congeneric species in general. The same was found, for a different region of mtDNA, for species of Macaca by Hayasaka et al. (1996).

This of course makes sense under molecular clock assumptions. Avise and Ball (1990) demonstrated by computer simulations that the longer the two populations have been separated, the more loci should differentiate them (although smaller effective population sizes will reduce the time it takes for them to become reciprocally monophyletic); if it is true that species have been separated
for longer than infraspecific populations, then one should expect that on average there should be higher genetic distances between them.

But scanning Ayala’s tabulations, one observes that the values of D for the four levels overlap widely, at least in his vertebrate examples; well-differentiated species pairs may actually differ less than do sibling species; species of any kind frequently differ less from each other than do subspecies; subspecies may differ less than do undifferentiated local populations. In Thorpe’s (1983) dataset, again, the overlap between levels is enormous. One would be foolhardy to insist that above a certain level two taxa are merely subspecies, while below it they are species.

On average, species of the same species-group differ by 9.55% of cytochrome b positions in rodents but only 6.83% in bats (Bradley and Baker 2001 p 963; Table 1) and 7.86% in macaques (Hayasaka et al. 1996 pp 1048–1049; Table 2, and some extra distances calculated from Table 1). But these are average differences; it is possible for different species to differ by much less than this—as little as 2.23% in the smallest of the rodent comparisons, 2.50% in bats, and 3.72% in macaques. The conclusion, and it is a very important one, is that we must simply adjust to the fact that well-differentiated species pairs may show extraordinarily little sequence divergence in a given DNA region.

### 18.2.5 Hybridizing species

Reproductive isolation may or may not be a result of speciation, and the old controversy over whether such RIMs as may exist precede the evolution of species-specific diagnostic states, or are a by-product of them, is unresolved. But it is evident that species that hybridize are not necessarily each other’s closest relatives: cases are known where sister species are intersterile, while less closely related species are not (Baum 1992).

Hybrid zones between distinct species are referred to as tension zones (Key 1968). Their width depends on migration, which acts to widen the zone, and selection, which acts to reduce its width (Barton and Hewitt 1985). The most stable hybrid zone is one which follows an ecotone; it widens and narrows in different regions and apparently tracks the ecotone over space and time (Arntzen and Wallis 1991). Whether it persists over long periods of time is arguable; can there be selection against individuals that lack a focused search image, such that they waste their reproductive effort by producing hybrids? Paterson (1978, 1980) argued strongly that such reinforcement is unlikely; Barton and Hewitt (1985), on the contrary, considered that reinforcement may in fact occur, although it is only one of a number of factors which may help to maintain the zone.
There are several well-analyzed cases of hybridization between perfectly distinct species. Two newts, *Triturus cristatus* and *T. marmoratus*, are broadly parapatric in western Europe, where they are characteristic of flat, open country and more hilly, wooded regions, respectively (Arntzen and Wallis 1991). They hybridize in places, and the hybridization appears to be one-way, as hybrids all possess *T. cristatus* mtDNA. The authors argued that there was selection against hybrids, both on theoretical grounds (chromosomal incompatibility) and because of the high frequency of digital anomalies among them; they did find, nonetheless, evidence that genes had leaked across the hybrid zone into the “pure” species on either side. Whereas for Key (1968), this would have been reason enough to combine the two species into one, Arntzen and Wallis (1991) argued against such an interpretation, on the grounds that the two essentially maintain their genetic integrity overall, and even in the hybrid zone itself only 4% of the population are F1 hybrids.

Among mammals, the classic study of a hybrid zone is that by Hunt and Selander (1973) of two commensal mouse species, *Mus musculus* and *M. domesticus*, across the width of the Jutland peninsula, Denmark. This hybridization process too is asymmetric; mtDNA from one species leaks some distance well beyond the hybrid zone proper into the other species, although interestingly this does not apply along the same species boundary in southern Germany, where the mtDNA and nDNA boundaries are concordant.

Species do not have to form actual hybrid zones: well-differentiated sympatric species may also hybridize. *Cercopithecus ascanius* and *C. mitis* hybridize sporadically in East African forests, and in Gombe National Park this is so common that the frequency of hybrids is comparable to that of the parent species (Detwiler 2002). But hybridization need not be quite as blatant as this: phenotypes may be for all intents and purposes unaffected, and only the foreign mtDNA remains as a kind of fossil of recent or more remote episodes, as in a case of two species of deer (*Odocoileus*) in Texas, which appear quite distinct but may possess the other species’ mtDNA (Ballinger et al. 1992).

A phenomenon occurring in many hybrid zones is the high frequency of rare or otherwise unknown alleles, termed hybrizymes (Woodruff 1989). They may result from recombination within loci, increased mutation, or relaxed selection. The very existence of hybrizymes suggests that, for whatever reason, new genetic features may arise in hybrid zones (or perhaps more probably they survive there, where they do not within the pure species), and it is extrapolating only slightly to suggest that there may be entire species of hybrid origin, analogous to those which have arisen by polyploidy so frequently among plants. A striking example has been deduced in mammals: the Stumptail or Bear Macaque, *Macaca arctoides*, has Y chromosome sequences typical of the *M. sinica* group but mtDNA
sequences similar to *M. fascicularis* (Tosi et al. 2000), and most likely the species is derived from a hybridization event. *M. arctoides* is very far from being morphologically intermediate between its presumed parental species and has obviously changed a good deal since it was first generated. Other examples of species that may be of hybrid origin are the Wisent or European Bison, *Bos bonasus* (Verkaar et al. 2004), and Père David’s Deer, *Elaphurus davidianus* (Meijaard and Groves 2004).

### 18.3 Speciation

#### 18.3.1 How does speciation happen?

The literature on speciation is considerable (Barton 2001), but one cannot improve much on the categorization by White (1978) into seven general modes:

1. **Strict allopatry without a population bottleneck.** This is ordinary vicariance on either side of a newly arisen barrier, when the two daughter populations are both large. If selection is not operating to push the two apart, they still have a low but finite probability of slowly diverging by stochastic processes as modeled by Kimura (1983).

2. **Strict allopatry with a narrow bottleneck of one component.** This is of course founder effect; typically the component that is not bottlenecked remains identical to the parent species, while the bottlenecked component diverges rapidly.

3. **Extinction of intermediate populations in a chain of races.** There is commonly some clinal variation through the range of a species and, if geographically and morphologically intermediate populations go extinct, the endpoints will already be somewhat differentiated. There is some fear that this process may be happening today as populations of widespread species vanish in all but a few isolated reserves.

4. **Clinal speciation.** A cline may become stepped over time, resulting in para-patric speciation.

5. **Area-effect speciation.** If incompatible alleles or gene complexes arise in different parts of a species’ range and spread, there may be selection against hybridization when they meet.

6. **Stasipatric speciation.** This was White’s own discovery (White 1978, 1979). A chromosomal rearrangement arises and spreads despite low heterozygote fertility and becomes established as soon as its frequency increases to the point where homozygotes begin to be generated. This process, as discussed
by Bush et al. (1977) and Meester (1988), creates new linkage groups and breaks up old ones, which might affect gene expression; it may be promoted under some forms of social organization such as clans and harems, by limited vagility and dispersal, distributional patchiness, individual territoriality, and generally by low effective population size.

7. **Sympatric speciation.** This is the one guaranteed to start fistfights in bars. Mayr (1963) argued strongly against it, but from time to time it has nonetheless surfaced as a real possibility. Maynard Smith (1966) argued that differential habitat assortment could lead to speciation in the absence of any geographic separation; Bush et al. (1977) urged that it might even be the dominant mode of speciation among parasites and parasitoids. More recent models have tended to invoke sexual selection (Higashi et al. 1999; Kondrashov and Kondrashov 1999).

Speciation, then, may occur by drift, natural selection, or sexual selection, with or without founder effect; it may involve intrinsic mechanisms, such as chromosomal change, or it may be simply a matter of divergence following allelic substitution; and it may be allopatric, parapatric, or sympatric. We may infer the relative importance of these from phylogenetic studies (Groves 1989; Barraclough et al. 1998), although care must be taken to ensure that only sister-groups are compared in such exercises.

### 18.3.2 How long does speciation take?

“Species flocks” in cichlid fish are monophyletic clades which are endemic to a single lake, such as Lake Victoria, Lakes Tanganyika, and Malawi in the Central African Rift, and Barombi Lake in Cameroon. In a very innovative study, McClune and Lovejoy (1998) used cytochrome *b* to compare differentiation, within species flocks, among sympatric species to that among allopatric sets. The actual sequence divergence between sister species of fish was 2–8%; this compares to that among geographic populations within given species, which was from 0% to 5.6%. The overlap between percent sequence divergence within species and that between species (allowing for the possibility that some species pairs were not sister species) was “time for speciation” (TFS); as the divergence associated with allopatric speciation was 2.0–5.6% and that associated with sympatric speciation was 0–1.25%, TFS was calculated to be 1–2.27 Myr in allopatry but anything from “instantaneous” to 0.77 Myr in sympathy (using a cytochrome *b* divergence rate of 2.5% per million years). They admitted that these times could well be overestimated because of incomplete lineage sorting.
(which in some Haplochromine cichlids “appears to be epidemic”) and possible hybridization. The monophyletic swarm in Lake Victoria consists of 300 species; as the lake is only 12,400 years old, they calculate about 1,500 years per speciation event.

That speciation may take a very short period of time indeed can be inferred from looking at changes that have occurred in taxa which have been introduced by human agency outside their natural ranges. Three examples have become especially well known:

- Mice (*M. domesticus*) introduced apparently by the Vikings to the Faroe, Shetland, and Orkney Islands differ noticeably from the ancestral Scandinavian and/or British populations (Davis 1983). The mice of two of the Faroe Islands were described as new subspecies by Degerbøl (1942), as was the hare whose history is similar.

- Australian wild rabbits differ appreciably from their parent population in the British Isles, and they are more and more different the further we go from Geelong (Victoria), the place where the small founding population was first introduced in 1859 (Taylor et al. 1977).

- In 1916, a pair of rock wallabies (*Petrogale penicillata*) was imported from New South Wales to Hawaii for a private zoo, but they escaped, and over 200 of their descendants now roam wild in the Koʻolau Range of Oahu. They are noticeably different in pelage, size, cranial characters, and body proportions from their ancestral population (Lazell 1980; Lazell et al. 1984).

These three examples are perhaps not actually speciation but certainly serve as examples of very rapid changes following the establishment of small founder populations. The Australian rabbit case also illustrates the principle of isolation-by-distance, meaning that populations that are more distant geographically may also be more distant genetically (the basis for clinal variation); considering the way rabbits disperse, not *en masse* but by forming small founder colonies, the cause of isolation-by-distance may, at least sometimes, be a progressive series of founder events.

Groves (1989 pp 44–46) pointed out that the amount of heritable change following a founder event will vary in magnitude. Presumably small gene frequency changes will be most frequent; the greater the degree of change, the less probable—but not vanishingly so. In the “sudden origins” model of Schwartz (1999), mutations of large potential effect, but recessive at first, accumulate in a population until their frequency reaches the point at which all at once they appear abundantly in homozygous state: similar, but not identical, to the way stasipatric speciation was envisaged by White (1978). If this is so, most founder events will
not result in speciation but some certainly will. The trick, of course, is to get these new species to survive when they come into sympatry with their parent species.

18.3.3 Is there phyletic speciation?

If founder effect is theoretically the most powerful way of generating new species, would speciation result if an entire population underwent the process (in other words, bottlenecking)? The “founder-flush” model of Carson (1975) relies on the fact that during the reexpansion phase after a population crash there will be no density-dependant selection until carrying capacity is reached, so all kinds of zany mutants will survive to form new and potentially fit recombinations. Petry (1982) argued that this can apply to entire species. There are in fact real-world examples of this; some were described and discussed by Groves (1989 pp 39–43). Consider: if this happened from time to time in a species with high visibility in the fossil record, it will mimic the appearance of anagenesis. We should be alert to the possibility of examples of punctuated equilibria without cladogenesis; there may be cases where rapid change in a fossil species occurs entirely in situ rather than in small peripheral isolates. Gould (2002 pp 779–780) was well aware of this; as he noted, cases of phyletic speciation might be distinguished from cases of allopatric speciation because the ancestral species is nowhere to be found.

18.3.4 Can we detect speciation modes in the fossil record: does anagenesis really exist?

Groves (1989) pointed out that, if speciation really does drive evolution, we would expect to find that highly autapomorphic species have fossil records characterized by high rates of speciation, whereas their more plesiomorphic sister species would have had many fewer speciation episodes. The human species is highly autapomorphic, and indeed the fossil record depicts rampant speciation almost throughout; unfortunately the case lacks a control, as the plesiomorphic sister-group, the chimpanzee, has no fossil record whatever.

The reality of anagenesis has been vigorously defended by White (1995), and it is true that there are plenty of cases in the Plio-Pleistocene record where a pair of time-successive species (White’s example is in the suid genus Kolpochoerus) seem to merge into each other when fossils from intervening time zones become better known. Yet, as White (1995) notes, there are big problems which get in the way of a proper interpretation of the record: gaps are always there, key fossils are frustratingly incomplete, and examples of the same presumed species are not
represented by comparable parts. A further problem is that “well-known cases” of anagenesis, which were plausible hypotheses in the state of knowledge of yesterday, may call for reanalysis when they are looked at anew in the twenty-first century. A classic case of anagenetic gradualism was the inexorable onward-and-upward evolution of the East African Plio-Pleistocene elephant *Elephas recki*. Recently, Todd (2005) has shown that all is not well with this icon of anagenesis, and a radical revision of the species (if it is one species) is in order.

How, in any case, would it work? Eldredge and Gould (1972) invited their colleagues to consider the implications of long-term sustained directional change—an environment exerting exactly the same selective pressures over perhaps millions of years. The most plausible scenario would be the Red Queen Effect of Van Valen (1973), in which species themselves are constantly changing—usually degrading—their own environment and are forced to evolve continuously in order to track these changes. The Red Queen model incorporates everything from African elephants knocking down trees, and so destroying their own future food supply, to evolutionary arms races between lions and antelopes: everything, in other words, in which no evolutionary stable strategy is resolved. (The African elephant scenario is actually cyclical, although it is likely that natural climatic change does reset the cycles each time; such cases will therefore be most likely episodic, not continuous and sustained, see Cameron and Groves 2004.)

In cases of sustained trends, there is an alternative: the effect hypothesis of Vrba (1980). In this model, new species are continually generated, and these are then subject to the forces of selection; the ones which survive are those whose innovations happen to make them fitter in their particular habitat. At each round of speciation this will occur, so that, if the environment has not changed much, an appearance of gradual directional change will result.

If it is difficult to test gradualism in the fossil record, how easy is it to test speciation? The answer is that, unless we catch it in the act, we can only infer it (from a pattern of stasis followed by sudden change); even then, the possibility of phyletic speciation (see above) cannot be ruled out. But there are a few cases where we have caught it in the act. The Morrisville cowpasture site, apparently documenting the evolution of a new species of trilobite, as described by Eldredge (1986, Chapter 3), may be one. Another, which appears to have survived all the criticism that was thrown at it, is the simultaneous eruption of new species in a dozen mollusk lineages in the Lake Turkana deposits (Williamson 1981). This latter case, as argued by Groves (1989), seems to flout the PE model because the new species were generated not by founder events in small isolates but in huge populations. Yet it might instead shed light on mechanisms, because it occurred not instantaneously, as in cases we have considered above, but over thousands of years, more as envisaged by Eldredge and Gould (1972) in their original paper;
it might thus be explicable by a Kimura-type mechanism rather than by a Schwartz-type one.

18.4 Conclusions

I have briefly surveyed the arguments over what is a species; the most nearly objective definition, the Phylogenetic Species Concept, relies on genetic discontinuities and stands at the base of the “evolutionary species,” which has a distinct evolutionary history and fate. How do species arise?

Gradualism (anagenetic change) is surprisingly hard to justify theoretically; if it could be demonstrated in the fossil record, this is something we would simply have to live with, but it is doubtful whether it can; at least one case (E. recki) that has been cited as a classic example of it has been found to need revision, and it is probably time to reexamine other supposed examples in fossil mammals such as the African pig lineages.

Speciation is, in most cases, most likely a type of rapid cladogenesis; there is evidence to suggest that in some instances it may be very rapid indeed. It may be allopatric, by a small population becoming isolated from the main body of the species, but stasipatric and even sympatric models are also plausible. Phyletic speciation, though it does not involve lineage splitting, also falls under the definition of speciation. In the fossil record, speciation would show up as a punctuation.

But, because a punctuation in most of the geological column—prior to the Neogene, anyway—is likely to take no more than “the duration of a bedding plane” (to quote Gould 2002 p 768), we are unlikely to catch it in the act very often. In most cases, we have to fall back on inferring its operation. For example, a period of stasis followed by the sudden appearance of a new species is best explained by PE. And, if indeed speciation is the real focus of evolutionary change, then one expects to find evidence of much lineage splitting in the fossil record of highly autapomorphic species, and very little in that of a plesiomorphic species; a proposition which is in need of testing in the future.

References


Ayala FJ (1975) Genetic differentiation during the speciation process. Evol Biol 8: 1–78
Hunt WG, Selander RK (1973) Biochemical genetics of hybridization in European house mice. Heredity 31: 11–33
Thorpe JP (1983) Enzyme variation, genetic distance and evolutionary divergence in relation to levels of taxonomic separation.
Tosi AJ, Morales JC, Melnick DJ (2000) Comparison of Y chromosome and mtDNA phylogenies leads to unique inferences of macaque evolutionary history. Mol Phyl Evol 17: 133–144
White MJD (1978) Modes of speciation. WH Freeman, San Francisco
Wright S (1968) Evolution and the genetics of populations. Chicago University Press, Chicago
19 Human Environmental Impact in the Paleolithic and Neolithic

Wolfgang Nentwig

Abstract

With increasing population size, the environmental impact of Paleolithic and Neolithic societies increased. The most apparent effect of hunters and gatherers was the disappearance of large herbivorous mammals (megafauna) soon after men appeared on a new continent (best examples are the Americas). The overkill hypothesis attributes this to increasing human population pressure and efficient hunting techniques, but also to effects of changing climate and vegetation. The loss of the Pleistocene megafauna probably led to modifications in vegetation structure, higher fire frequency, and a reduction in biodiversity. With further increasing population density, man became a food producer through the invention of agriculture and stock farming. In Mesopotamia since 6,000 years BP, the development of irrigation systems led to a considerable increase in the agricultural area, but since the fourth millennium BP large area losses due to irrigation mistakes (salinization) are reported. Generally, deforestation and erosion processes are known from several regions of the Mediterranean area since 7,000 BP.

19.1 Introduction

The environmental impact of a human population is usually a function of its population size. Therefore, when considering the possible environmental impacts of ancient man, the first question concerns the number of humans who were living at the end of the last glaciation, roughly some 12,000 years ago. Of course, no census was made at that time but some simple calculations already provide us with figures which are probably quite reliable.

With the colonization of Australia and the Americas, H. sapiens doubled its territory to 136 million km². Roughly, one-fourth of this area was habitable, the rest being deserts, polar regions, high mountains, etc. These people made their
livings as hunters and gatherers and their density was low. In human hunter-gatherer populations living in modern times, we know comparable societal structures that give an overall population density of 1/10 km². This estimate leads to a total world population size of 3.4 million for *H. sapiens* around 12,000 years BP.

An assessment of the environmental impact of these Paleolithic societies can only address the game the hunters killed. Possible effects could include an overexploitation of their resources and thereby modifications of the influence these animals had on their environment. More precisely, game could have become rare or even extinct, thus their grazing effect could have been reduced and the vegetation structure changed, e.g., from grassland to a woodland ecosystem. One major topic of the environmental impact of early man therefore concerns his impact on large game animals.

After most habitable areas of the world had been colonized around 12,000 years BP, a further increase was only possible through an increase in population density. Since the hunting and gathering mode of life did not allow a higher population density, a change of lifestyle was induced by population growth beyond the Paleolithic density. Between the Mesolithic and Neolithic, agriculture and stock farming were independently invented several times and provided a new source of food: food gatherers became food producers. This “Neolithic revolution” allowed a considerable population increase. Although around 7,000 years BP the area under mixed farming was too small to have much effect on global population size, there was a constant increase in the cultivated area and population size. According to McEvedy and Jones (1978), the global population size increased by 50% in the seventh millennium BP and by 100% in each of the following three millennia. About 3,000 years BP, the doubling time dropped to 500 years and the growth speed reached its peak ([Figure 19.1](#)). This corresponds to a world population of 100 million at 2,500 years BP and a population density of 3/km². Regionally and locally, population density was much higher and we can easily imagine multiple environmental impacts arising from these Mesolithic to Neolithic societies.

In comparison: today (2005), the world population comprises nearly 6,500 million and the average density is 48/km². Environmental damages, such as reduction of biotic diversity and destruction of ecosystems, are present everywhere and constitute a serious danger to future human living conditions (Nentwig 2005). To understand the current situation and facilitate the implementation of sustainable solutions, it is essential to analyze the historical record. In this context, it is useful to go back to the roots of the environmental impact between the Paleolithic and the Neolithic.
The Paleolithic period as considered here was characterized by a dramatic climate change from the last glacial stage to a warm period which persists until today. In large areas of Europe and Asia the glaciers slowly but constantly receded as the warming progressed. Landscapes formerly occupied by ice and snow now became covered by vegetation and developed with the further succession into ecosystems with higher biodiversity and productivity. So, not only did the area habitable for humans increase but also its quality. People obviously pressed forward into these newly developing habitats in a migration direction principally from south to north and used the resources that they found or developed there. Among these resources, large mammals were their main nutritional resource.

One remarkable exception to this general postglacial migration pattern concerned the Americas. Throughout their history, humans had never reached the New Worlds because the sea of the Bering Strait or the enormous glaciers on Beringia blocked the access to North America. During the last glaciation, an enormous amount of water was bound in the world’s ice cover and the sea level was 100 m lower than it is today. The Bering Strait was dry and formed the Beringia land bridge. It was covered by open grassland and had a rather mild climate. This area was settled by humans, but the huge ice shield covering North America prevented further expansion. Only the melting of the ice mass allowed free access to the New World when the coastal route first appeared (probably 13,500 years BP) and later a land corridor across the continent (probably 11,000
years BP) (Dixon 2001). Possibly, the door was only open for a short while because soon the melting water from the glaciers raised the sea level and the Beringia land bridge disappeared under the sea, but it was the starting point for the very fast colonization of the Americas. In America, the hunters met a very diverse association of large herbivorous mammals that had never been in contact with man before, and they hunted them very effectively.

In both parts of the world considered here, America and Eurasia, fast spreading groups of hunters relied on large herbivorous mammals and their predators, together the so-called megafauna. The hunting success of Paleolithic man was so enormous that most species of the megafaunal complex were over-exploited and became extinct. This was the first strong environmental impact in human history, resulting not only in a considerable loss of very important species but also in a change of the environment, because the extinct mammals had influenced and directed the development of this environment as key species (Nentwig et al. 2004).

These historic events are still controversial, since it is not easy to imagine that the Paleolithic humans were able to kill giant mammals such as mammoths. But we will see that their hunting techniques were very well developed. Another reason for the ongoing debate concerns the simultaneous occurrence of warming up, ecosystem succession, appearance of man, and disappearance of the megafauna. The striking coincidence between these parameters is a major support for the overkill hypothesis but has also been used to argue in favor of the alternative ecological hypothesis. Finally, we will have a look at the differences between the otherwise similar events in America and Eurasia.

19.2.1 The megafaunal complex

The term “megafauna” does not refer to a taxonomic unit but is weakly defined by size. In the Late Pleistocene, the megafauna consisted mainly of large herbivorous mammals: in Xenarthra the armadillos and sloth order, the 6 m long Eremotherium and Megatherium ground sloths, or the turtletile protected giant Glyptodon are well known. Among rodents, the Neotropic giant beaver Castoroides and giant capybaras should be mentioned. Notungulata, such as the hippopotamus-like Toxodon, reached 3 m in length. Among Perissodactyla, many horse and tapir species occurred on all continents but rhinoceros were restricted to the Old World. The Artiodactyla comprised many hog, pig and peccary species, Hippopotamus, several camel and lama species, a few giraffe and elk species, many deer, pronghorn and antelope species, several bison species, many buffalo, goat and sheep species. Among Proboscidea, several species of mastodons,
mammoths, elephants, and gomphotheres were distributed throughout most continents and represented the most impressive members of the ancient megafauna. There were also predators among the megafauna, such as some dog species, cats, bears, and hyenas. Several species of sabertooth tigers were probably specialized on large and slow-moving prey.

In the Pleistocene epoch, these species formed characteristic species complexes, well adapted to their specific ecosystems. It is striking that the Late Pleistocene faunas had higher numbers of species than their modern counterparts. Such a higher diversity can perhaps be attributed to a much higher degree of specialization and perhaps even coevolution than many large mammals exhibit today (Graham 1979). The population density of the larger species of the megafauna was low; they lived long and had a low number of offspring. An annual population loss of 10%, which Pleistocene hunters usually achieved, could have become crucial for these animals within a few dozen years. Many of these species were rather specialized to their environment, e.g., fed only on a narrow spectrum of plants (McDonald 1979). A suitable comparison may be provided by the community of large herbivores in the East African savannahs and their well-defined grazing specialization.

A typical Late Pleistocene grazer community of the arctic steppe, tundra, or open and patchy parkland/forest consisted of mammoth (Mammuthus) and rhinoceros species in the Old World, horse (Equus) and camel (Camelops) in the New World, plus deer species including the giant Megaloceros giganteus, bison (Bison), as well as goat and antelope species.

### 19.2.2 Paleolithic hunting techniques

*H. sapiens* evolved in Africa, a continent which is still today famous for the richness and abundance of its large mammals. The social structure of a human clan and their competence in organizing themselves into hunting groups enabled them to become perfect hunters. While hunting, humans constructed a variety of weapons and improved them for given purposes. Early weapons, such as lances and spears, however, were only suitable for very short distances up to 10–15 m. These weapons nevertheless enabled man to kill the largest mammals such as mammoths or rhinos.

The first long-range weapon was the spear thrower or atlatl (Garrod 1955), usually made of wood or antler. It consists of a shaft, 30–100 cm long, and a spear. One end of the shaft is held in the hand while the spear is placed in a surface groove at the other end or fixed by a hook or a loop. This hook keeps the spear in place while it is held in a slightly raised position by the same hand that holds the
handle end of the atlatl. By the leverage of this catapult-like start, the hunter throws the spear with an acceleration and force up to three times more powerful than that of a hand-thrown spear. Spear throwers could wound or kill animals up to distances of about 30 m.

The oldest spear thrower was found in Northwest Africa and is over 25,000 years old. The oldest European record from the Magdalenian is 17,000 years old. This weapon was used in North America at least 10,000–13,000 years ago, and it is probable that the invading people already possessed it (Dixon 2001). The typical stone projectile tips of atlatl spears were found associated with the remains of mammoth, extinct bison species, and other large game animals. People of New Guinea also used spear throwers and the Australian aborigines called it woomera.

In Europe, the spear thrower was replaced by the bow and arrow some 12,000 years BP. One of the oldest pictures of a bow was discovered in the Grotte des Fadets (Dept. Vienne, France), and its age has been estimated as 12,000–13,000 years. The oldest European records of bows and arrows are at least 10,000 years old; the corresponding North American records are usually considered to be much younger (4,000 years), but Dixon (2001) mentions the introduction of the bow and arrow into North America at about 10,500 BP. The bow allowed the use of rather small projectiles, many of which could be carried when hunting. While the spear thrower required a movement of the hunter which could warn the game, the bow was a hidden weapon. It could be used over larger distances and experienced hunters killed animals from up to 50 m away (Figure 19.2).

19.2.3 The megafaunal overkill

The combined effects of changing climate and vegetation, human population pressure, appearance and colonization of new habitats, and the possession of suitable weapons led, on some continents, to a strong decimation of the megafauna complex, to the point of extinction of many large mammals. This striking coincidence between the expansion of man and the extinction of species suggests that humans killed so many individuals of given species that population densities dropped below a critical value, region by region, and they finally became extinct. The interaction of such disparate agents and the short time of the event provoked Martin (1966, 1984) to call it “overkill.” He also used the term “blitzkrieg” to describe the scenes at the migration “front,” where the first invaders into North America met the megafauna. The attributes of this blitzkrieg are a rapid movement of the human population into an area not previously inhabited, the
possession of a megafauna-hunting technology, and the virtual simultaneous extinction of megafauna species due to direct hunting.

In contrast to Martin’s overkill hypothesis, climate and environmental changes have also been blamed for the disappearance of species (climate hypothesis, ecological hypothesis). In their original forms, these hypotheses assume that the megafauna was not capable of adapting to the changing environmental situation and became extinct. But many species survived earlier climate changes, repopulated lost territory from relict areas, and proved to be astonishingly flexible. These hypotheses also fail to explain why large mammals were more prone to extinction than smaller species, or why extinction in the Americas was more severe than in Africa or Europe. They give no explanation for the lack of synchronization between events on different continents and the lack of comparable cases of extinction during previous glacial periods. Therefore, the ecological theory has to include man as an important factor. Perhaps the population size of the megafauna had been reduced due to climate and vegetation change, habitat loss, and worsening of diet quality. Human predation, however, was a key factor which finally led to the extinction of many megafauna species (Guthrie 1984; Owen-Smith 1987; Barnosky et al. 2004).

It is widely accepted that species usually become extinct for a variety of reasons or combinations of parameters. The overkill hypothesis only deals with the colonization process of man and the implied interference with the
megafauna. These events happened in Africa in a less sharply defined period during the last 100,000 years. Eurasia had been populated in several waves during the last 50,000 years (and of course also much earlier). Australia was invaded 50,000 years BP, America 13,500 years BP, Madagascar 1,500 years BP, and New Zealand 1,200 years BP.

Most of the American megafauna had no enemies and probably did not display specific flight behaviors. When this part of the world was settled by man for the first time, these animals had no experience with people and did not know what a danger man was to them. In other areas, such as Africa and Eurasia, the Upper Paleolithic hunters and megafauna had already coexisted for some time. Even if man could easily prey on these giants with their superior weapons and hunting techniques, they did not eradicate them completely. The time of appearance of man is obviously important: the later man arrived in an area, or the longer the isolation was, the more sudden and complete the extinction of the megafauna was (Figure 19.3). The larger an animal was, the higher the risk to be hunted by man: the relative number of extinct species per size class of terrestrial mammals decreased with decreasing body size, but only a few aquatic species and no climbing species became extinct (Martin 1984).

Africa is the area of origin of *H. sapiens* and its megafauna escaped larger incidences of Late Pleistocene extinction. Only a few species disappeared during that period due to human hunting (e.g., giant buffalos). The loss amounted to

![Figure 19.3](image_url)

Schematic diagram showing the decrease of the megafauna as percent survival of large animals on three continents and two large islands. After Martin (1984)
14% of all genera larger than 44-kg body weight. The usual explanation is the long
time man and game lived together and got used to each other. As an additional
explanation, Lambrecht (1980) mentions that tsetse flies (and perhaps other
pathogens and parasites) limited the range of man and the amount to which
they could use the continent, e.g., as pastoralists, so that human population
density always remained lower in Africa than in other continents.

Several times in its Pleistocene history Europe lost large mammal species, but
most of these survived in Africa, Asia or, in the case of the musk ox, even in
America, and later repopulated Europe. Only three European genera completely
disappeared in the Late Pleistocene: the giant deer *Megaloceros*, the woolly rhino
*Coelodonta*, and the woolly mammoth *Mammuthus*. The situation in Asia is
much less well investigated but may be similar to Europe or Africa.

Australia was the first isolated area in the world which, due to climate change
and subsequent lowered sea level, could be invaded by man within a short period
of time. This happened some 40,000–60,000 years ago and exposed the very
specific Australian fauna to human influence. A total of 86% of all genera with
species larger than 44-kg body weight disappeared during the following millen-
nia. The extinct megafauna included marsupials, such as the rhino-sized *Dipro-
todon* (estimated body size 2,000 kg), giant wombats, the tapirlike *Palorchestes*,
and several large kangaroo species, but also the giant varanid lizard *Megalania*
(larger than the Komodo dragon), the giant horned tortoise *Meiolania*, and the
ostrichlike thunderbird *Genyornis* (Dromornithidae).

It was in the Americas that the loss of the megafauna was most obvious. In
North America, 74% of all genera with species larger than 44-kg body weight
became extinct; in South America 80%. All mammoths and gomphotheres
(Proboscidea) (*Figure 19.4*), all horses (Equidae), all rhinos (Rhinocerotidae),
and all camels (Camelidae) disappeared. Among Xenarthra all ground sloths,
such as the 6-m long *Eremotherium* and *Megatherium* or the 3-m long giant
armadillo *Glyptodon*, disappeared. The youngest remains were usually about
11,000 years old, shortly after the time when the first hunters had arrived.
Among the surviving large mammals was one group of Eurasian origin, late
arrivals to North America (moose, wapiti, caribou, musk ox, grizzly bears, and
some others), which probably was already conditioned to man. Another group of
survivors consists of South American relatives of North American extinct species
(e.g., guanaco, tapir, capybara, peccary, and some others). Perhaps hunting was
less effective in the dense tropical forests of South America where most of these
species are found today.

Madagascar became settled by man only 1,500 years BP. Within a few cen-
turies, 7 of 17 primate genera completely disappeared, all of them large and
diurnal, while the small and nocturnal species survived. A pygmy hippopotamus,
the endemic order Bibymalagasia and two species of giant land tortoise also became extinct. All 6–12 species (depending on taxonomy) of elephant birds (Aepyornithidae) became extinct, the largest measuring more than 3 m in height. Comparable to Madagascar, New Zealand was reached by humans only recently (1,200 years ago). Its very special fauna lacked mammals and the moas (Dinornithidae), huge birds up to 3 m in height, probably partly filled this niche. All moas, 13–29 species (depending on taxonomy) died out due both to extensive burning of their habitats and to hunting. In Madagascar and in New Zealand, both islands lost most or all megafauna species within 500 years after the first people arrived.

19.2.4 The black hole after the megafauna loss?

Killing the megaherbivores not only led to the loss of these species but also all effects which these animals exerted on their environment, especially the vegetation, ceased or became altered. All species directly or indirectly depending on the megaherbivores, also became extinct. So by killing the megaherbivores, early man directly and indirectly reduced the biodiversity. This reduced the functional services of the biodiversity and altered the structure of his environment.

The most obvious effect concerned the structure of the vegetation. The open nature of the Late Pleistocene forests was probably the result of the large
herbivorous mammals which prevented, by their grazing activity, a succession toward denser vegetation. Ungulates and carnivores could use mammoth paths through thickets. The loss of such open space certainly reduced the habitat quality for many smaller species.

The mammoth steppe was a large, productive grass-dominated steppe reaching from Alaska to Eurasia, an ice-free refugium during the last glaciation with mammoths, steppe bison, horses, woolly rhinos, musk oxen, sheep, camels, and Saiga antelopes. Extinction of a few key species caused by human hunters, as the keystone herbivore hypothesis of Owen-Smith (1987) describes, led to a reduction of soil disturbance and nutrient input. This resulted in the replacement of nutrient-rich grasses by nutrient-poor mosses and sedges, and formed a moss-dominated tundra (Catling 2001). According to a model proposed by Zimov et al. (1995), the well-balanced equilibrium between these two ecosystem types was disturbed by the human removal of megaherbivores (Figure 19.5).

Such grazers also contribute by seed predation to the distribution of seeds, thus facilitating the spread and germination of their most important food plants. Janzen and Martin (1982) compared the Pleistocene Central American situation
to tropical, forested African habitats, where elephants, rhinoceroses, zebras, and other large herbivores are key factors for the open landscape, its mosaic structure, and seed dispersal. In Pleistocene America, ground sloths, elephant-like gomphotheres, horses, and other large herbivores also had this function. Many tree species produce large amounts of fruits with thick shells which could only be opened and eaten by these strong herbivores. The indigestible seeds are released in the dung and show high germination rates. Since between the time of eating and defecating the animal moves, these seedlings grow up far away from their parent plants. This guarantees better germination conditions than under the parent tree where up to 100% of all seeds are destroyed by seed beetles or mould. Seed dispersal by large mammals also leads to a higher diversity within the plant association, indicating that many of those plants that are rare today must have been common before.

Without large herbivores, the vegetation becomes less mosaiclike, more biomass remains uneaten and dries up, and fires are more frequent and intensive. In comparable African ecosystems, elephants prevent the savannah from becoming dense shrub thickets and increase its patchiness. The Pleistocene megafauna probably had similar fire-preventing effects and enabled a higher biodiversity.

19.3 Agriculture

With increasing human population density, man had to switch from being a food gatherer to becoming a food producer. This transition occurred between the Mesolithic and the Neolithic, when agriculture and stock farming were invented several times independently. The most ideal prerequisites were fertile soils in a plain or hilly landscape, if possible, close to a river and a mild climate. A sufficient water provision had to be guaranteed, either by natural precipitation (rain-fed agriculture), or by the water table, or by artificial irrigation (irrigation agriculture). In consequence, larger proportions of the human population became sessile, managed to maintain arable fields by clearing the vegetation and by building dams and digging ditches in a cooperative effort. Economy and social order developed. Important further steps toward modern civilization included the use of metal, invention of writing, urban agglomerations, and state systems.

The major areas where agriculture first developed were in the Near East (in the Nile valley in Egypt, in Anatolia, northern Syria, in the Jordan valley, in Mesopotamia, and in valleys of the Zagros Mountains of Iran), India (the Indus Valley), and China (along the lower Huang Ho/Yellow River). The first
agricultural systems started to develop in the Near East about 12,000–10,000 years BP, in India and China at least 7,000 years BP.

A cultivated soil loses fertility when no nutrients are applied (e.g., by fertilizers). Thus early agricultural systems used shifting cultivation and had a rather high land turnover. Since arable fields replaced natural habitats, a potential ecological impact arose from the loss of the natural habitat and from a lack of knowledge or possibilities being applied to maintain the arable fields. Unsuitable agricultural practice causes, for example, an accumulation of salt in the soil (salinization) or a loss of humus (by erosion). Erosion is also provoked by large-scale deforestations, e.g., in order to obtain timber or firewood for cooking and metal processing.

19.3.1 Salinization: the example of Mesopotamia

Around 7,000–8,000 years BP, farmers from the northern mountainous areas settled the broad plain between Euphrates and Tigris (Mesopotamia). In the southern parts of Mesopotamia, rainfall was insufficient and water from the rivers had to be used to irrigate the fields. Later this area became known as Akkad in the north and Sumer in the southern part of this river system.

About 6,000 years BP, the Sumerians began forming small agricultural villages. Later, they increased their settlement density and invested much work into canals and dikes to make use of the annual flood which spread a thick layer of fertile silt across the flood plain. This labor needed thousands of workers and a strong coordination. The planning of the canal network and the distribution of irrigated land and water also demanded political control. Between 5,500 and 5,000 years BP, the Sumerian societal organization had developed into temple communities, some of which later became the rich city-states of Sumer, of which the best known is Ur. These communities were capable of organizing the irrigation management of large parts of their territory for many centuries.

In the early dynastic period (starting 4,900 years BP), the patterns of settlement and canals were considerably rearranged to better utilize the irrigation water. A general decline of the water volume of the large rivers forced the Sumerians to use the water as economically as possible. Simultaneously, they increased the irrigated area to face the increasing demand due to population growth. In the third dynasty of Ur (2,112–2,004 BC), the highly sophisticated irrigation systems retained the water as long as possible and prevented surface runoff into the canals and rivers as much as possible (Klengel 1989; Nissen 1999). However, this forced the farmers to give up an older field fallow rotation system and to cultivate their fields permanently. The old fallow rotation had allowed the
water table to fall after the harvest, but permanent cultivation of the same field now became problematic.

The semiarid climate and the high evapotranspiration of plants and soils led to a considerable water loss into the atmosphere. The dissolved salt accumulated in the soil and in the near-surface water table, which had hardly any lateral movement to carry it away. In consequence, the groundwater became more and more saline. Since irrigation water was reduced to an absolute minimum, no surplus water washed the surface salt into the river. At this advanced stage, even excessive irrigation could not solve the problem because the wet soil additionally raised the water table by capillary forces and dissolved salts were brought into the root zones of the plants.

Freshwater on average contains 3.5 g salt/L. Many plants need up to 10,000 m$^3$ water/ha in one season. Without sufficient surface runoff, this salt load accumulates to 35 t/ha (≈3.5 kg/m$^2$). After a few years, such a soil is so heavily salt-polluted that agriculture is no longer possible. Especially in arid or semiarid environments, soil salinization is one of the most severe hazards in modern agriculture (Nentwig 2005).

The increasing extent of salinization and its threat to the Sumerians can be seen from the decreasing amount of (salt-sensitive) wheat (emmer) that was grown in the saline fields and its replacement by (salt-tolerant) barley. But even barley did not do well, since the overall yield per ha decreased within 800 years to one-third of the initial yield (Figure 19.6). Around 3,700 years BP, wheat cultivation had to be abandoned in several parts of southern Mesopotamia (Jacobsen and Adams 1958; but see Powell 1985). At that time, the Atrahasis
Epic complains “The black fields became white, that broad plain was choked with salt” (Lambert and Millard 1969).

During the third millennium BC, the Sumerian empire managed to compensate for the increasing loss of local agricultural productivity, e.g., in the southern part of the province Lagas, by shipping in surplus yield from other provinces. The Sumerians knew of the salinization phenomenon, and they had an evaluation system to estimate the yield of a field according to its degree of salinization (Pettinato 1967). There are also reports on the salinity status of individual fields, showing increasing salinization over several centuries (Jacobsen and Adams 1958).

The Sumerian environmental crisis, here revealed as the salinization crisis, certainly had several causes. In the fluctuating climate of the period following the last glaciation, the period from 5,000 to 3,700 years BP is characterized as a warm period. This influenced the precipitation in the surrounding mountains, the water charge of the rivers, and evaporation of water from soils. On the other hand, there was an increasing demand for arable fields and irrigation water because the population steadily increased. The response of the government to this increasing gap between yield and demand of agricultural products was the complete control of the irrigation systems and of the total agricultural production. The preserved documents (thousands of written reports, such as storage lists of agricultural products) demonstrate that this centrally planned economy basically became a mere administration of a deficit. When foreign peoples, especially the nomadic Amurritians, who, probably due to the increasing dryness in their eastern steppe areas, left their homeland and increasingly attacked the Sumerian empire; it was already so weak that it could not persist for long. Those parts of the empire which depended on food supplies from the north could no longer be sufficiently provided for. The resulting famines enfeebled these areas and they were overrun by the conquerors (Nissen 1999). In the next centuries, a couple of smaller political entities replaced the Sumerian empire, the population of southern Mesopotamia decreased, and larger cities developed much more in the north. Thus it is too simple to state that the Sumerians disappeared due to the salinization of their fields, but it is beyond question that the increasing soil salinity played an important part in the breakdown of their culture.

19.3.2 Deforestation and erosion

Already the early development of agriculture was characterized, due to population growth, by an increasing demand for arable fields. People cut the natural
vegetation, performed some tillage operations, and used a field for a few years. On hillsides, rainfall washed unprotected soil down into the rivers below. Due to loss of soil and due to lack of fertilizer, the fertility of such fields usually declined after a few years. Greater areas had to be cultivated and steeper land was used, both factors increasing erosion.

In the Mediterranean area, the destructive effect of human activities has been documented since 7,000 years BP, at least locally (Figure 19.7). Later cultures increased land use pressure over wide areas. Especially, the Greeks and the Romans completely destroyed large forests. This caused catastrophic soil erosion and prevented a sustainable agriculture (Wagner-Hasel 1988; Runnels 1995). Plato (427–347 BC) was already complaining about the soil erosion after the forests had been cut down.

Figure 19.7
Due to deforestation and agriculture on the surrounding mountains, the river Menderes (also known as Meander) in SW Turkey transported huge amounts of soil into the Latmian Gulf of the Aegean Sea. The upper third of the Gulf was filled in 5,500–3,000 years BP, in the following 400 years the central third, and since 2,600 BP, the lowest third. The inset shows the estimated erosion load with maxima in the Neolithic and in the Hellenistic period. Combined after Bay (1999)

Apart from the demands of agriculture, deforestation and subsequent erosion also increased because of the rising demand for wood as firewood or timber. In the case of Mesopotamia, timber (e.g., for roofs of palaces and temples or for ships) had to be transported from the surrounding mountains via the Tigris and
Euphrates into the cities. The oldest sailing boat is known from a model in a Sumerian grave dated about 5,500 years BP.

Already 9,000 years BP, Anatolian artisans discovered how to use copper. Since copper is a rather soft metal, its use in tools and weapons was limited until bronze could be produced. The first records of bronze, an alloy of copper and tin, date from 5,500 to 5,000 years BP (Egypt, Crete, China). At that time, copper and tin were found only in a few locations and demanded long trading routes (tin from Britain into the Mediterranean area), whereas iron was much more common. By about 3,400 years BP the Hittites in Anatolia knew how to process iron. All this handling of metal produced a high energy demand, which in those times could only be provided by charcoal. Therefore, the first metal processing occurred in forested areas (e.g., Anatolia) and enhanced deforestation.

### 19.3.3 Influence on vegetation succession: case study of Central Europe

After the last glaciation, large areas of Europe were again covered with forest. Since the ice-free refugia had been far away in the southeastern and southwestern parts of Europe, the colonization speed was different for many species. In central Europe at 10,000 years BP, pines and birch dominated (*Pinus, Betula*), at 8,900 years BP hazel-rich (*Corylus avellana*) mixed oak (*Quercus*) forests developed (with alder *Alnus*, ash *Fraxinus*, elm *Ulmus*, lime *Tilia*), after 3,800 years BP also with beech (*Fagus sylvatica*) and hornbeam (*Carpinus betulus*).

At the same time man colonized these areas, first as a hunter and gatherer, but soon after as a farmer. Through collecting seeds and fruits, the gatherers distributed them as well, as if they were concerned to have as many useful shrubs and trees around as possible. Hazelnut is the most obvious tree which could readily spread, but oak and beech also could. The speed and extent of the colonization of *Fagus sylvatica* shows an astonishing similarity to the spread of the Neolithic culture in Europe. In addition to a direct spread as a result of the collection and distribution of beechnuts, also indirectly, by cutting down of less-favored trees, natural competition with *Fagus* became reduced (Lang 1994). Similar effects are probable for other shrubs and trees with useful fruits or nuts. Such human-influenced distributions in Europe resulted in a postglacial colonization of Central Europe that was much faster than by natural means (Portnoy and Wilson 1993) and that differed from a natural vegetation composition.

Around 4,000 years BP, most pollen profiles from Central Europe already show cereal pollen, i.e., farmers had spread over this area, cut the forest down,
and cultivated the land. Since these early agricultures could only use a field for a few years, this was followed by a longer period of fallow, and sometimes the field was also used as pasture. Some abandoned fields became covered with shrubs, and occasionally they developed into an early successional forest, until they were cut down or burnt again for agricultural use. Even if arable fields comprised only a low percentage of the total area, the human influence on the forest composition and structure was much greater because of this immense land turnover rate.

Also grazing had an important effect on the forest structure. Herds of swine and cattle grazed in the forest and prevented natural regeneration. In the long term, open forest-landscapes developed. In combination with land use for agriculture, the main effect of the Neolithic farmers was an opening of the landscape. Many other plants and animals profited directly and indirectly from this. The opening of the woodlands and their conversion into an open landscape slowly increased the biodiversity in Central Europe. This increase started already at the end of the glaciation but became intensified from 7,000 years BP on, when Neolithic man took his first steps toward the cultural landscape we have today.

19.4 Conclusions

Population growth was the driving force between Paleolithic and Neolithic for humans to spread over the world and to increase their inhabited area. Already as hunters they had overexploited their game resources and eradicated many mega-fauna species. As farmers, their overexploitation of the environment continued: the intensive use of irrigated fields increased salinization problems, while deforestation for agriculture or for wood supply caused erosion. In all cases, the resulting environmental damage persists until today: extinction is forever; salinization is one of the most serious threats to modern agriculture in a semiarid environment, and nowadays the Mediterranean basin is approaching more and more desertlike conditions.

Thus, Mesolithic to Neolithic humans were already capable of exerting a serious environmental impact. There is no difference from the damage of today in quality, only in quantity. This conclusion contrasts with the expectations many modern people have. We are still imprinted by Rousseau’s cliché of the noble savage and sympathize with the last primitive people of the globe. Traditionally, we admire the Greek and Roman cultures and their ancestors. Generally, we have no precise information on their way of living but implicitly assume that these primitive or ancient cultures lived in balance with nature. This wishful thinking
can easily be explained by our daily experience of the human environmental impact worldwide, but nevertheless this assumption is wrong.

Acknowledgment

This chapter profited very much from the profound knowledge of Thomas Naumann (Siegen) to whom I am very grateful. My thanks for valuable discussion and advice go also to Felix Kienast (Birmensdorf), Brigitta Ammann (Bern), and Cecily Klingler (Bremgarten).

References


Lang G (1994) Quartäre Vegetationsgeschichte Europas. Fischer, Jena


20 The Dentition of American Indians: Evolutionary Results and Demographic Implications Following Colonization from Siberia

Christy G. Turner II · G. Richard Scott

Abstract

This chapter uses dental morphology to make inferences about how the New World was first colonized. The major emphasis is on the initial Macro-Indian migration based on dental traits observed in Paleo-Indian, Archaic, and more recent prehistoric crania. The major results are: (1) Arctic and Subarctic native dentitions differ enough from those of Macro-Indians to indicate separate migrations. (2) Clustered MMD values show three Macro-Indian branches of North Americans, South Americans, and mixed North and South. (3) There is no marked branching depth for these three dental divisions, which fits the hypothesis of a single rapid Paleo-Indian colonization event. (4) The minimally divergent North and South American dental divisions are most likely the microevolutionary result of dispersal-dependent population structure and lineage effects. (5) No genetic bottlenecks can be identified at Panama. (6) The small amount of New World internal dental divergence favors colonization of South America soon after the settlement of North America. (7) There are no obvious clines, frequency trends, or geographic groupings for individual dental traits. This suggests little or no selection and that after leaving Siberia, population size increased sufficiently to limit genetic drift. (8) There is no sign of any Old World or Oceanic dental pattern other than Northeast Asian Sinodonty. All things considered, including New World and Siberian linguistics, archeology, genetics, route considerations, and relevant natural history, dental analysis supports the Late Pleistocene Ice-free Corridor, Clovis or epi-Clovis settlement hypothesis, and the Greenberg Amerind or Macro-Indian language evolution model.
20.1 Introduction

This chapter reviews the dental morphological evidence for the pre-Columbian colonization of the New World. This evidence, in conjunction with genetic information, is used to challenge colonization scenarios based on measurements of prehistoric skulls. There are two main anthropological questions for which answers have been sought using the variation of New World dental morphology and comparative information from Old World and Pacific basin populations. First, how many colonizing migrations can be postulated on the basis of New World dental variation? To answer this question, analysts have tried to determine if there are any dental correspondences with modern or ancient New World culture areas (yes), environmental types (no; however, Jantz et al. 1992 found that anthropometric variation was correlated with geography), or linguistic families (yes) (Greenberg et al. 1986; Scott and Turner 1997). These findings suggest that New World dental variation is more closely linked to ethnicity (i.e., migration) than to environment (natural selection or physiological adaptation). Roughly put, dental variation is strongly caused by nature, whereas cranial variation has a significant nurture component (Kohn [1991 p 273] notes that “… environmental factors contribute a significant amount to the observed variability in craniofacial morphology and growth.”). Because of linguistic and ethnic correspondences with dentition, Arctic and western Subarctic dental variation is most parsimoniously interpreted as due to two migrations of people dentally differentiated to a slight degree from an earlier migration (Clovis or epi-Clovis). All three inferred migrant groups possessed minor variants of the Sinodont dental pattern, and as such, are closely related and share a common Northeast Asian ancestor sometime in the relatively recent past. Presumably, each originated in different geogenetic regions of northeastern Asia (Turner 1985) after separating from the common ancestral stock that presumably lived further to the south, say central China. As will be shown, the earliest colonizers must have given rise to all modern North and South American populations except for Arctic and Subarctic groups. We refer to this first migrant group as Macro-Indian or Amerind, linguistic names (Ruhlen 2000), and Paleo-Indian or Clovis, synonyms used by archeologists (Haynes 1987). The term epi-Clovis has been recently proposed to denote the Late Pleistocene ancestral cultural condition in Beringia prior to the development of the fluted points that are the hallmark of the Clovis big-game hunters (Turner 2002). Another name used in Alaska to denote epi-Clovis is Nenana complex (Hoffecker et al. 1993). In addition to these just-cited articles, other major reviews of the archeological evidence for the peopling of the Americas can be found in Carlson (1991), Fiedel (2004), Haynes (2002a, which contains an immense bibliography, 2002b), Haynes (n.d.), Hoffecker and Elias...
(2003), Madsen (2004), Meltzer (1993), Rogers et al. (1992), West (1996), Yesner (2001), Yesner and Pearson (2002), and many others. We emphasize the importance of archeology and archeologically derived teeth since both provide direct diachronic epigenetic evidence for human biocultural prehistory.

In the Old World, the Sinodont dental pattern occurs in modern and prehistoric Northeast Asians of China, Tibet, Mongolia, post-Jomon Japan, Korea, and eastern Siberia (Hanihara 1968; Zoubov and Haldeneva 1979; Hanihara 1991; Scott and Turner 1997). In contrast, Late Pleistocene and recent peoples of mainland and island Southeast Asia (Thailand, Indonesia, Borneo, Philippines, Taiwan, Jomon Japan, etc.), as well as Polynesians, Micronesians, and early Sri Lankans, possess a dental pattern called Sundadonty, which is simpler and retains more of the character of a hypothesized earlier pattern (Proto-Sundadonty) than does the more specialized Sinodonty (Turner 1983, 1990a; Scott and Turner 1997). Modern and recent Europeans, as well as the earlier Cro-Magnons, possess a simplified dental pattern, which is similar to that found in modern India, north Africa, and western and central Asia (Irish 1993; Haeussler 1996; Hawkey 1998). The simplest of modern human dental patterns is found in Europe. It more closely resembles the dental pattern of modern Africans and Australians than Sinodonty.

The second question involves controversial archeological and craniometric evidence claims proposing that the New World was initially colonized by a pre-Clovis, Sundadont population that was less Mongoloid cranially than modern Indians. These claims presume an earlier entry into the New World than the classic Clovis-first model. Claims for pre-Clovis range widely, from 15,000 to more than 50,000 years BP. We also address this question.

To put the peopling of the New World into a broad framework of dental pattern evolution, we follow a phylogenetic reconstruction that seems to have the best general concordance with other lines of evidence derived from archeology, linguistics, genetics, and natural history. By dental pattern we mean a standardized set of independent discrete crown and root morphological traits whose occurrence (presence or absence) and expression (small, moderate, large, very large; shape variation, etc.) in one major geographic region, such as northeast Asia, differ considerably from what is found in other major natural regions, such as Europe. The authentication of these patterns has been based on repeated dental sampling of the geographic regions, which produces trait frequencies like those originally used to define the regional patterns. On dental grounds we exclude prehistoric Europeans and Africans as having anything directly to do with the peopling of the New World until the Christian era. Similarly, teeth indicate that peoples of Oceania and Southeast Asia played no role in the colonization of the New World. Long before Siberians reached Alaska, human
populations had evolved their major modern dental patterns. Scott and Turner (1997) identify these as belonging to five groups: Africa south of the Sahara, western Eurasia (Europe, Middle East, north Africa, India), Sahul-Pacific (Australia, New Guinea, Melanesia), Sunda-Pacific (mainland and insular Southeast Asia, Polynesia, Micronesia), and Sino-American (northeast Asia and the Americas). Finally, we exclude as impossible all claims for an autochthonous creation of Native Americans, as based on some religious beliefs, as well as on very flawed reasoning (Tyler 1998).

To put the peopling of the New World into a broader human evolutionary context, we should first consider what we know about human evolution in Asia during the Upper Pleistocene. Sometime before 50,000 BP, there were anatomically modern humans living in Southeast Asia, although only the south China Liujiang skull may be this old or older (Pope 1992 p 275). However, the 50,000 BP date approximates the earliest archeological appearance of human-made stone artifacts in Australia (Roberts et al. 1994). All human skeletal remains found in Australia, some dating as far back in time as 30,000 BP, are anatomically modern. Australia could only have been colonized from Southeast Asia, and at a few locations along the route, only with the aid of some form of watercraft. We hypothesize that the early Southeast Asians and their Australian colonists possessed a dental pattern similar to what is found in modern Australians. Turner (1992c) has called this hypothetical pattern Proto-Sundadonty, which was estimated by averaging the dental trait frequencies obtained from archeologically and ethnographically derived skeletal samples collected in several localities throughout Australia. In their early terrestrial and oceanic dispersal eastward out of Southeast Asia, Proto-Sundadonts, in time, became ethnic Melanesians and Australians. Their dental pattern is much like that found in modern and recent Southeast Asians as well as modern sub-Saharan Africans suggesting Proto-Sundadonty and the African dental pattern were derived from an as yet unidentified common ancestral pattern. Before the end of the Pleistocene 10,000 years ago, Southeast Asian teeth had evolved into the slightly more specialized Sundadont pattern as part of the general drift toward dental complexity in eastern Eurasia, in contrast to the Late Pleistocene drift toward increased simplicity in western Eurasia. The areal extent of Sundadonty increased with the northward coastal expansion of these modern humans into Taiwan, and Jomon Japan, where Sundadonty remains to the present day in unadmixed Taiwan aborigines and the Ainu of northern Japan. In the more interior part of the East Asian mainland, probably in north China, Sundadonty evolved further into the even more complex and specialized Sinodont pattern. In time, nomadic groups possessing Sinodonty expanded into Mongolia, eastern Siberia, and eventually the New
World. In the later Holocene, Southeast Asians traveled the Pacific by large sailing watercraft to become island populations of Polynesian and Micronesians, retaining their Sundadont pattern. Thus, all around the Pacific basin there are three distinctive dental morphological patterns: Proto-Sundadonty (Sahulland: Australia and Melanesia), Sundadonty (Sundaland: mainland and island Southeast Asia; island East Asia), and Sinodonty (Sino-America). At least 50,000 years are involved in the evolution of this Asian-Pacific dental patterning. In terms of time involving only several thousands of years, these patterns are evolutionarily quite stable and remarkably consistent in ethnically related groups living in strongly dissimilar ecological settings. Such demonstrable stability is part of what makes dental morphology eminently useful for evaluating long-term diachronic and synchronic affinities, population origins, and microevolutionary processes.

20.2 New World Arctic and Subarctic dental variation

While all New World teeth conform with the Sinodont dental pattern, first defined on the basis of prehistoric Chinese teeth, those of Arctic and Subarctic Eskimos, Aleuts, and speakers of Na-Dene languages differ slightly from those of Macro-Indians (Scott and Turner 1997, n.d.). For example, pronounced upper incisor shoveling is less frequent in Eskimos and Aleuts than it is in Macro-Indians. It tends to be intermediate in Subarctic Indians. On the other hand, Eskimos and Aleuts have greater frequencies of three-rooted lower first molars than do Macro-Indians. Again, the western Subarctic tribes are intermediate. When the three New World groups—(1) American Arctic, (2) Northwest North America, and (3) North and South America—were graphed for mean trait frequencies and standard errors of 23 crown and root traits, 10 of the traits show the intermediate nature of Northwest North America viz. Eskimo-Aleuts and American Indians in general while 11 traits showed similar frequencies across the 3 groups (Scott and Turner 1997). The only two traits that failed to exhibit intermediary or similarity were incisor interruption grooves and odontomes, but the differences were not dramatic in either instance.

Multivariate and univariate analyses and comparisons show clearly that all New World groups are dentally more like Northeast Asians than like any other major geographic group in the world. Nevertheless, the teeth of New World Arctic and Subarctic peoples are generally more like those of Northeast Asians than are the Macro-Indian populations, which as archeological evidence suggests, means that the Macro-Indian colonization occurred before or separated from that of the
modern Arctic and Subarctic populations. It has been inferred that the concordances between New World archeological groups and biological populations likely are: Big-game hunting Clovis fluted point makers = Paleo-Indian Sinodonts; smaller game-hunting Paleo-Arctic biface and microblade makers = Na-Dene/Greater Northwest Coast Sinodonts; Maritime Anangula blade makers = Aleut-Eskimo Sinodonts. What holds these concordances together is their individual geographic distributions for artifact types, dental divisions, and language families.

20.3 Macro-Indian dental variation

Comparative analyses of Native American dentitions by Haeussler and Turner (1992), Hanahara (1979), Perzigian (1984), Powell (1993), Scott (1973, 1991), Scott et al. (1983, 1988), Turner (1983, 1985, 1986, 1990a, 1992a, 1992b, 1993, 2002), as well as contemporary genetic studies by Callegari-Jacques et al. (1993), Cavalli-Sforza et al. (1988, 1994), Horai et al. (1991), Laughlin and Harper (1988), Stone and Stoneking (1993), Szathmary (1979, 1993), Torroni et al. (1992), others reviewed herein, and most osteological studies by Alekseev (1979), Kozintsev (1995), Steele and Powell (1992), and Brace et al. (2001), are in good agreement that the ancestors of Native Americans originated in northeast Asia. However, a few osteological studies are not in full accord with this general inference. Howells’ (1989) worldwide craniometric analyses showed male Peruvians and Californians clustering with males from Europe, while his male Arikara Indians clustered with male New Zealand Moriori. Neves and Pucciarelli (1991) argued on multivariate grounds that early South American crania are morphometrically closer to Australians even though their principal components plots show the early South Americans equally similar to East Asians and Europeans. Inasmuch as archeology (West 1981, 1990; Dillehay and Meltzer 1991; Stanford and Day 1992; Meltzer 1993), linguistics (Greenberg et al. 1986; Greenberg 1990; Greenberg and Ruhlen 1992), and paleo- and neogeography (Hopkins 1982; Laukhin 1997), as well as dentition, genetics, and most osteology, strongly support a Northeast Asian origin for all Native Americans, the two cited cranio-metric studies are notable exceptions to the Northeast Asian-American anthropological and other correspondences as well as the Sino-American dental correspondences (Scott and Turner 1997). Nevertheless, in recent years there has been a concerted effort to bring back to life the notion of a pre-Clovis occupation of the Americas (Hall 2000).
20.3.1 A brief history of the pre-Clovis revival as it bears on dental variation

There has long been disagreement between Americanist scholars who line up with the late Aleš Hrdlička (1925) by arguing that the colonization of the New World was late compared to the amount of human evolution in the Old World, in contrast to scholars who believe there is evidence for much older human occupation in the New World. For example, Krieger (1964) (preprojectile occupation), Carter (1957) (very ancient occupation), Müller-Beck (1967) (Mousteroid element in Americas), and others posit very early migrations (>20,000 BP) of peoples from the Old to the New World. There have been any number of archeological claims for pre-Clovis occupation of the New World, but as Meltzer (1993) humorously noted, the claims usually have a “shelf life” of about five years. Among modern physical anthropologists, several revivalists of the pre-Clovis, and/or non-Mongoloid origins include Neves and Pucciarelli (1991, and several similar articles), Crawford (1992), Lahr and Haydenblit (1995), and Steele and Powell (1992).

Lahr and Haydenblit (1995) proposed that Sundadonty was present in a series of South American crania from Patagonia and Tierra del Fuego. In a lengthy review article, Lahr (1995) proposed that either Sinodonty evolved in parallel in Asia and the Americas, or there were two migrations ancestral to Native Americans, i.e., a Sundadont group followed by a Sinodont group. These proposals are inconsistent with our South American dental observations. No where has the senior author found any South American series or individual skull that could be considered as having the Sundadont pattern and this includes the Lagoa Santa remains housed in Brazil and those curated in Denmark, as well as Archaic samples from coastal Brazil, Chile, Ecuador, and Peru. We suggest that the observations of Lahr and Haydenblit on the degree of trait expression were impacted by dental attrition. Wear most likely caused the underscoring of crown traits that led them to incorrectly propose the presence of Sundadonty (Burnett et al. 1998; Turner et al. n.d.). Tooth wear might also have misled Powell and his associates in their observations on small Paleo-Indian samples.

When Steele and Powell (1997) evaluated the craniometric data of two ancient skulls from Nevada (Spirit Cave, Pyramid Lake) that dated ca. 9,500 BP, they found that the two skulls failed to cluster with any of the 22 comparative modern populations. Despite this result, they felt that the skulls were more closely aligned with south Asian, Pacific, and Australian populations than with North Asian and recent North American Indians. They concluded that “the studied Paleoindians arrived in the Americas prior to the establishment of the crania.
shape that is distinctive of recent Northern Asians and North American Indians, and that the colonization of the Americas was more complex than has previously been proposed” (Steele and Powell 1997 p 218).

Taken altogether, the form of the Paleo-Indian skull relative to most of the comparative samples is not only a reflection of geogenetic linkages but also sedentism and its related nutritional, growth, health, and activity benefits and stresses. Thus the Africans cluster together (geogenetic linkage) despite both nomadism and sedentism being represented. The same can be said for Europe, northeastern Asia, and the Oceanic set. Given that archeological remains of Paleo-Indians strongly suggest a nomadic hunting life way, then the cranial differences between Paleo-Indians and recent Native Americans should have been interpreted along economic lines as much as geogenetic. Only in the last sentence of their article do Steele and Powell (1992) remark on the possibility of “adaptational” factors contributing to the cranial differences between Paleo-Indians and recent Native Americans. All but one sentence in this article is clearly aimed at identifying possible “genomic” differences between Paleo-Indians and recent Indians.

It should be noted that Steele and Powell (1992 p 329) speak of Paleo-Indian crania as not being “...classically sinodont in craniofacial appearance. Instead, it differed by appearing as much like modern southern Asians [recall they use this term to refer to Chinese, who have traditionally been classified as Mongoloids] as it did recent North American Indians and northern Asians. In this respect our findings resemble the contentions of previous scholars that the earliest recovered samples were proto-Caucasoid or proto-Mongoloid.”

As far as we know, there has never been any analysis carried out to show what sort of relationship there is between craniofacial morphology and Sinodonty or Sundadonty. Steele and Powell have assumed that vaults, faces, and teeth go together like a hand in a glove. In fact, Sinodonty is found in people whose craniofacial variation includes every shape in South and North America including very robust California Indians, heavy and long-skulled Southwest US Basket-makers, and very robust long-headed Archaic crania from Mexico, Brazil, and Chile, to very round-headed gracile Southwest Puebloans, Chinese, Japanese, Buriats, and a long list of other long- and round-headed Northeast Asians. As for early Native Americans having been proto-Caucasoids, a notion perhaps most strongly championed by Birdsell (1951), Harris and Turner (1974) showed that dental morphology was strongly in opposition to such typological thinking (Birdsell was a student of Hooton whose 1930 book The Indians of Pecos Pueblo was the ultimate in, but not the end of, typological thinking. The crania of Pecos Pueblo had such a wide variety of cranial forms that Hooton’s typological classification of that within-group variation led Gladwin (1947) to publish his
hilarious spoof on anthropology titled *Men Out of Asia*. Gladwin used Hooton’s cranial classification to postulate multiple distinct migrations into the New World. The first to enter were coast-following plant-eating “Australoids.” Next came meat-eating Clovis hunters, presumably of a Negroid strain. Algonquins entered third, Eskimos fourth, and Mongoloids, fifth. All this, of course, is a jumbled mix of typology, morphological dating, age-area concept, and primitive–modern scaling.

### 20.4 Genetics and the peopling of the New World

During the first half of the twentieth century, scholars who investigated issues relating to the settlement of the New World came primarily from archeology and skeletal biology. During the past 50 years, however, genetic data have played an increasingly important role in debates on Native American population history. Researchers have applied three different kinds of genetic data to the problem. Initially, focus was on nuclear genetic markers, from red blood cell antigens and serum proteins to red cell enzymes, immunoglobulins, and white blood cell antigens (cf. Mourant 1954; Mourant et al. 1976; Roychoudhury and Nei 1988; Cavalli-Sforza et al. 1994). In the 1980s, a new and different type of genetic marker was used, one derived from the single-stranded DNA of the mitochondria (mtDNA) which was transmitted, with but few exceptions, through maternal lineages (for reviews, see Cann 1988; Long 1993). Finally, in the 1990s, the Y chromosome was sequenced and a surprising variety of interesting polymorphisms emerged (cf. Hammer and Zegura 2002; Jobling et al. 2004). The literature in the field has expanded at an exponential rate during the past two decades. It is beyond the scope of this review to provide a detailed synthesis of genetic studies. Our goal is to highlight some of the results from the three different sorts of genetic data to determine the extent to which the analyses of synchronous genetic data corresponds with peopling models based on diachronic dental data.

#### 20.4.1 Nuclear markers

Around 1950, the nascent field of human genetics started to weigh in on the issue of Native American origins. On that date, W.C. Boyd published his seminal work *Genetics and the Races of Man*. Therein, he utilized available data on three red cell antigen systems (ABO, Rh, MN) to set up a classification of modern human groups. He deemed Native Americans to be sufficiently distinct from other groups to warrant their own racial category. In this regard, he noted that Indians
lacked the “r” allele of the Rh system and the A^2 and B alleles of the ABO system, although Eskimos did have a low frequency of B. Indians also showed a relatively high frequency of the M allele and a marked dichotomy in frequencies of the A and O alleles of the ABO system. Populations south of the US–Mexico border showed essentially 100% O alleles while some North American groups, in particular Algonquians, Athapaskans, and Eskimo–Aleuts, exhibited high frequencies of the A allele. While similarities to Asian populations were noted (lack of r and A^2), Boyd felt that the high frequency of the M allele and the very low frequency of B was sufficient to distinguish Asian and American “races.”

Laughlin (1951) used blood group data along with anthropometric and osteometric comparisons to assess the affinities of Aleut populations. He felt the presence of the B allele indicated Aleuts were closer to Eskimos than Indians. Moreover, he noted that the “B present in the Eskimos is an indication of their recent Asiatic heritage . . .” (Laughlin 1951 p 119). Throughout his career, Laughlin (1963, 1966) adhered to the notion that there were two major groups in the Americas—American Indians and Aleut–Eskimos. After comparing Native Americans to Siberians for data on the ABO and MN systems, Laughlin (1966 p 473) opined that the “essential affinity of the Eskimo–Aleut stock with Asiatic Mongoloids, rather than with American Indians, is well attested.”

With the development of starch gel electrophoresis in the early 1950s, there was a dramatic increase in the number of genetic surveys across the Americas and throughout the world and in the number of genetic systems that became standard markers for population profiles. By the 1970s, the original three RBC antigen standards were complemented by the addition of many more antigen systems and serum proteins (e.g., Diego, P, Duffy, Kell, Kidd, Haptoglobin, Transferrin, Albumin, etc.). By the time of their massive worldwide synthesis, Cavalli-Sforza et al. (1994) were able to tabulate data on 120 nuclear alleles.

Beginning in the 1970s, Szathmary (1979, 1981, 1993 and elsewhere) played an important role in making inferences on long-term population history based on the analysis of Native American gene frequency profiles. One of her early efforts, prepared with Nancy Ossenberg, had the eye-catching title “Are the Biological Differences Between Eskimos and North American Indians Truly Profound?” (Szathmary and Ossenberg 1978). This paper developed the position that Indians and Eskimos were not as distinct as many scholars had presumed (cf. Laughlin 1963, 1966), even suggesting that the two groups might have differentiated from a common stock after their arrival in the New World. Until this time, authors had disagreed on the number of possible migrations to the New World but the consensus was that the ancestors of Aleut–Eskimos constituted not only a separate migration but the last major migration across the Bering Strait. While the arguments and analysis in these works are interesting, Szathmary limited her
comparisons between Aleut-Eskimos and northern Indians of the American Subarctic, to wit Athapaskans in the west and Algonquians in the east. We now know that these northern groups are much more similar to Eskimo-Aleuts than Indian groups from Mesoamerica, Central America, and South America (cf. Schanfield 1992; Cavalli-Sforza et al. 1994).

Authors who focus on GM allotypes have proposed different migration scenarios to account for variation in immunoglobulin genetic variants. Williams et al. (1985) observed that Gm$^{1,21}$ was present in “Paleo-Indians,” Na-Dene and Eskimo-Aleut samples with Gm$^{1,2,21}$ absent in Eskimo-Aleuts, and Gm$^{1,11,13}$ absent in Paleo-Indians. Na-Dene speakers had all three allotypes. This observation led the authors to conclude that Gm data supported the three-wave model of Greenberg et al. (1986) for peopling of the New World.

Also focusing on immunoglobulins, Schanfield (1992) pointed out that the work of both Szathmary (1993) and Williams et al. (1985) provided only a limited picture given their emphasis on North American native populations. When Schanfield synthesized Gm data from both North and South American populations, he found that South Americans differed consistently from non-Na-Dene North American Indians, Na-Dene speakers, and Eskimo-Aleuts. He feels his data suggest that “in the peopling of the New World, at least four separate migrant groups crossed Beringia at various times” (Schanfield 1992 p 381). He further suggests that the ancestors of South American Indians arrived before 17,000 BP while North American Indians arrived when the ice-free corridor opened up at the end of the Pleistocene. Both Eskimo-Aleuts and Na-Dene groups are thought to be later Holocene arrivals.

A grand synthesis of patterns of variation in nuclear genetic markers was accomplished by Cavalli-Sforza et al. (1994) in *The History and Geography of Human Genes*. This worldwide analysis of 120 markers in over 40 genetic systems devoted entire chapters to populations in each major geographic region, including a chapter on the Americas. To maximize the availability of genetic data across as many loci as possible, the authors combined data sets and came up with 23 New World groups, defined primarily on linguistic grounds. Importantly, the analysis included several Eskimo, Na-Dene, Siberian, North American, and South American groupings. After conducting a variety of distance analyses and two- and three-dimensional ordinations, Cavalli-Sforza et al. (1994 pp 340–341) concluded:

- The genetic patterns in the Americas fully confirm the three waves of migration suggested by dental and linguistic evidence: Amerinds, Na-Dene, and Eskimo. Their order in time is strongly suggested by their north-south geographical order. Further refinements may reveal that more than one entry contributed to...
the first wave, but the archeological information is contradictory and our understanding of the genetic pattern of Amerinds is incomplete, so that further investigations are required to settle this problem.

### 20.4.2 Mitochondrial DNA

When geneticists first discovered that mutationally induced variation in mitochondrial DNA (mtDNA) might help unravel human evolutionary history, there were severe sampling limitations. To harvest enough mtDNA for analysis, researchers had to collect human placentas, often a formidable task. The development of PCR (polymerase chain reaction) techniques in the early 1990s allowed researchers to obtain mtDNA samples from many kinds of tissue samples, including bone. This development revolutionized the study of mtDNA and research teams quickly pursued historical questions on every continent, among both living and earlier human populations.

Although mtDNA variation has been evaluated in groups throughout the world, the geographic region that has received an inordinate amount of attention is the New World. Because of sampling limitations, the early studies were regional in scope (Ward et al. 1991, 1993; Shields et al. 1992, 1993). With improved methods, expanding sample sizes, and cooperation among research teams, groups eventually addressed the bigger issues regarding the internal differentiation of New World populations and dispersal dates from ancestral Asian populations (Wallace and Torroni 1992; Torroni et al. 1993, 1994; Forster et al. 1996; Malhi et al. 2002; Eshleman et al. 2003).

Mitochondrial DNA variation is studied through a combination of restriction fragment length polymorphisms (RFLPs) and direct nucleotide sequencing of the relatively short hypervariable control region (HVR-I polymorphisms). Haplogroups are distinguished by a combination of RFLPs and HVR-I polymorphisms (Schurr 2004a).

Jobling et al. (2004 p 291) show the worldwide distribution of the major mtDNA clades, or haplogroups. Of the 27 major clades shown (lettered A to Z), New World populations exhibit the presence of only five haplogroups—A, B, C, D, and X. Numerous articles have been devoted to the issues of: (1) how these haplogroups vary within and among Native American populations; (2) whether or not they were brought by separate founding groups or differentiated after arrival in the New World; and (3) the time depth for the origin of each haplogroup.

Regarding mtDNA haplogroup variation, North, Central, and South American Indians (Amerinds), all exhibit haplogroups A, B, C, and D. Eskimo-Aleuts, by
contrast, have essentially no B and very little C (Schurr 2004a). Athapaskans were initially thought to lack the B haplogroup (Lorenz and Smith 1994), but it now appears to be present but infrequent in Na-Dene-speaking groups. Haplogroup X is limited to northern North American Indians from the Northwest Coast and Subarctic culture areas (Brown et al. 1998).

While the pattern of mtDNA haplogroup variation is coming into focus, opinions on what this variation means have yet to reach a consensus. Some early studies favored the notion that the haplogroups indicated four separate migrations into the New World (Schurr et al. 1990; Horai et al. 1993; Lorenz and Smith 1996). Torroni et al. (1994) concluded there were three migrations: two Amerind migrations, with an early dispersal of A, C, and D and a later migration that involved the B haplogroup. Athapaskans, with an exceptionally high frequency of the A haplogroups, were thought to constitute a third migration. Starikovskaya et al. (1998) also believed there was an early migration that carried the A, C, and D haplogroups with a later migration bringing B. Although there are several proponents of multiple migrations, the majority of researchers contend that the mitochondrial DNA variation evident among all Native Americans is best explained by a single migration event with haplogroup differentiation occurring after arrival in the Americas (Merriwether 1995, 2002; Kolman et al. 1996; Merriwether et al. 1996; Bonatto and Salzano 1997; Malhi et al. 2002; Silva et al. 2002).

From the outset, geneticists have argued that mtDNA is useful for not only describing patterns of variation but also for estimating times of divergence on a branching tree. Using either coalescence or distance methods of estimation in conjunction with several assumed mutation rates, most researchers addressing the issue of the initial New World settlement have arrived at very old dates for this event (Schurr 2004b). Several authors contend the first wave of migrants arrived in the Americas more than 30,000 years ago (Bonatto and Salzano 1997; Starikovskaya et al. 1998). Other workers give dispersal estimates of between 20,000 and 30,000 BP (Torroni et al. 1994; Silva et al. 2002). Starikovskaya et al. (1998) dated early Amerinds at 34,000–26,000 BP with a later migration at 16,000–13,000 BP. Torroni et al. (1994) tried to come up with an estimate that either agreed with an early entry date (30,000+ BP) or a late entry date (ca. 13,000 BP) compatible with the Clovis-first model. Instead of coming down on one side or the other of this debate, their estimate fell in the middle with a range of 22,000–29,000 years ago.

Despite differing opinions on the numbers of migrations and times of dispersal, mtDNA geneticists are in fundamental agreement that Native American haplogroups are of Siberian origin. Torroni et al. (1994 p 1162) reflect the sentiments of many when they say “We accept that all significant human entry
into the Americas was by way of Siberia during periods of glaciation, when a land bridge connected Siberia and the extreme northwest of the Americas.” When haplogroup X was discovered in American Indians, this was initially thought to represent a possible migration from Europe, or at least some founding Amerinds had Caucasian ancestry (Brown et al. 1998). If so, why is there no haplogroup H in Native Americans, as this is far more common in Europeans (ca. 40%) than haplogroup X (ca. 2%) (Schurr 2004a). An alternative to a direct transAtlantic migration of X-bearing Europeans was developed by Reidla et al. (2003) who feel that haplogroup X represents a fifth rare mtDNA clade that came into the Americas across Eurasia, originating as far west as the Near East (Reidla et al. 2003). Finally, there are no haplogroups that link modern or ancient Americans to Southeast Asian or south Pacific populations, a population source for early Americans favored by some osteologists (Neves and Pucciarelli 1991; Steele and Powell 1992, 1997; Powell and Neves 1999).

20.4.3 Y chromosome

The latest rage in genetic studies that focuses on the settlement of the Americas comes from the analysis of recently discovered polymorphisms on the Y chromosome. The long neglected Y, whose strict paternal transmission nicely complements the maternal transmission of mitochondrial DNA, exhibits two major types of polymorphisms on the nonrecombining segment (NRY) of the chromosome. First, there are point mutations that result in single nucleotide polymorphisms (SNPs). The mutation rates for these polymorphisms are slower than for the second type of polymorphism—short tandem repeats (STRs). Taken together, researchers use SNPs and STRs to define a diverse array of haplogroups and haplotypes. As with mtDNA, Y chromosome polymorphisms have been used to estimate the number of migrations, the timing of dispersal of populations to the Americas, and the Old World sources for this peopling event.

Evolving and diverse nomenclatural systems applied to Y polymorphisms make the literature on this system difficult to decode for nonspecialists. To allay confusion, workers convened a consortium in 2001 to develop a common set of standardized terms for Y chromosome binary haplogroups (Y Chromosome Consortium 2002; Hammer and Zegura 2002). Based on this new system, it appears that most Native Americans exhibit haplogroup Q, an observation that led some workers to conclude that a single founding male lineage is sufficient to account for the ancestry of all Native Americans (Tarazona-Santos and Santos 2002; Jobling et al. 2004). Haplogroup Q is, however, defined on the basis of several distinct SNPs and STRs, which other workers have interpreted as
indications of multiple migrations from Asia. Lell et al. (2002) suggest that the Q-M3 and P-M45a Y haplogroups dispersed eastward from central Siberia along with the C and D haplogroups of mtDNA. Moreover, these authors feel that the haplogroups P-M45b and RSP4Y came from south Siberia along with a subgroup of the A haplogroup of mtDNA (control-region sequence variant 16192T and the RsaI polymorphism at np16392). Karafet et al. (1999) and Ruiz-Linares et al. (1999) also conclude there were at least two founding Y haplogroups in the Americas.

Regarding the timing of dispersal, most estimates for arrival in the New World favor the long chronology (cf. Schurr 2004b). For their two migrations, Lell et al. (2002) estimate the first wave arrived in the New World ca. 20,000–30,000 BP with the second wave arriving much later, ca. 7,000–9,500 years ago. In contrast to those who see great time depth for Y haplogroups in the Americas, Seielstad et al. (2003) attempted to set an upper limit for initial entry into the New World based on a mutation (M242) that occurred just before the arrival of populations in the New World. Using a mutation rate of 0.18% per generation, the authors estimated that M242 arose around 15,000 BP. These workers, who consider 18,000 an upper limit for New World settlement, thus favor an entry into the New World that followed, rather than preceded, the Last Glacial Maximum.

20.5 Macro-Indian dental variation

Now that we have briefly reviewed contrasting models for the peopling of the New World from skeletal biology and genetics, we present in more detail the case that can be made from tooth crown and root morphology. Table 20.1 lists the archeologically derived Indian series used in this chapter. Altogether 3,584 individuals are represented in 17 North American groups and 11 South American groups. All the data in Table 20.1 were collected by the senior author. A key set of 29 largely independent, normal, age- and sex-free crown and root traits were scored for occurrence and expression using the Arizona State University dental anthropology system (Turner et al. 1991). The overwhelming numbers of individuals are pre-Columbian. Hence, there is effectively no chance for European, African, or Oceanic admixture in this Panamerican assemblage, whose chronometric ages range from protohistoric to late Paleo-Indian. Although the exact modes of inheritance are still under investigation for these and other dental traits, it is believed that each has a substantial genetic component for occurrence and expression (Scott 1973; Harris 1977; Berry 1978; Scott and Turner 1988, 1997; Nichol 1990).
Table 20.1

Frequency variation in archeological New World dental samples

<table>
<thead>
<tr>
<th>Trait, tooth</th>
<th>Shovel UI1 2-7/0-7</th>
<th>Uto-Aztec UP1 1/0-1</th>
<th>Enamel ext. UM1 1-3/0-3</th>
<th>One-root UP1 1/1-3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Break point</td>
<td>% No. %</td>
<td>% No. %</td>
<td>% No. %</td>
<td>% No. %</td>
</tr>
<tr>
<td>Sample</td>
<td>% No. %</td>
<td>% No. %</td>
<td>% No. %</td>
<td>% No. %</td>
</tr>
<tr>
<td>North America</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Archaic Canada</td>
<td>86.9 38 0.0 34</td>
<td>12.9 84 78.7 80</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iroquois Canada</td>
<td>71.7 39 0.0 116</td>
<td>26.0 231 75.4 207</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maryland</td>
<td>93.7 48 0.0 54</td>
<td>35.8 106 81.9 94</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Athapaskan</td>
<td>70.0 10 – 0</td>
<td>33.3 66 89.2 83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>California</td>
<td>97.7 88 1.1 91</td>
<td>41.7 213 83.9 211</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grand Gulch, Utah</td>
<td>100.0 7 0.0 33</td>
<td>45.9 48 94.1 51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chelly and Kayenta</td>
<td>97.6 42 1.7 59</td>
<td>61.4 96 97.4 76</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chevez Pass</td>
<td>71.4 7 4.2 24</td>
<td>55.1 29 83.9 31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>New Mexico</td>
<td>87.5 120 0.8 128</td>
<td>45.9 170 91.5 141</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grasshopper</td>
<td>91.0 133 4.0 124</td>
<td>49.1 157 84.0 81</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pt. Pines early</td>
<td>87.5 24 2.6 38</td>
<td>58.8 51 85.0 40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pt. Pines late</td>
<td>97.0 33 0.0 44</td>
<td>49.2 61 85.5 55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arkansas</td>
<td>87.8 49 4.1 97</td>
<td>33.1 127 88.8 125</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alabama</td>
<td>97.3 146 1.9 159</td>
<td>34.0 203 77.8 108</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coahuila, Mex.</td>
<td>75.0 8 6.5 31</td>
<td>21.1 76 87.3 63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cuicuilco and Tehuacan</td>
<td>94.7 38 1.7 59</td>
<td>33.3 66 91.2 57</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tlatelolco</td>
<td>100.0 39 0.0 85</td>
<td>50.8 122 91.0 122</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No.</td>
<td>797/869 20/1176 743/1906</td>
<td>1386/1625</td>
<td></td>
<td></td>
</tr>
<tr>
<td>South America</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panama</td>
<td>94.1 51 0.0 43</td>
<td>52.3 67 92.7 55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ayalan, Ecuador</td>
<td>90.2 51 1.4 74</td>
<td>44.3 106 88.7 97</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cotocollao, Ec.</td>
<td>100.0 28 3.7 27</td>
<td>54.0 37 89.3 28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chanduy, Ec.</td>
<td>88.2 17 0.0 12</td>
<td>40.0 20 94.7 19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Santa Elena, Ecuador</td>
<td>100.0 31 0.0 25</td>
<td>45.0 20 69.6 23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paloma, Peru</td>
<td>86.5 52 0.0 29</td>
<td>46.2 78 88.1 101</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Location</td>
<td>Mean</td>
<td>Range</td>
<td>Mean</td>
<td>Range</td>
</tr>
<tr>
<td>----------</td>
<td>------</td>
<td>-------</td>
<td>------</td>
<td>-------</td>
</tr>
<tr>
<td>Brazil 1</td>
<td>92.2</td>
<td>70.0–100.0</td>
<td>91.9</td>
<td>86.5–100.0</td>
</tr>
<tr>
<td>Brazil 2</td>
<td>95.0</td>
<td>92.2–100.0</td>
<td>91.4</td>
<td>86.5–95.6</td>
</tr>
<tr>
<td>Corondo, Brazil</td>
<td>95.0</td>
<td>92.2–100.0</td>
<td>91.4</td>
<td>86.5–95.6</td>
</tr>
<tr>
<td>Lagoa Sta, Brazil</td>
<td>91.4</td>
<td>86.5–95.6</td>
<td>91.1</td>
<td>86.5–95.6</td>
</tr>
<tr>
<td>Sambaqui, Brazil</td>
<td>91.1</td>
<td>86.5–95.6</td>
<td>91.1</td>
<td>86.5–95.6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Location</th>
<th>Mean</th>
<th>Range</th>
<th>Mean</th>
<th>Range</th>
<th>NA/SA $\chi^2$ 1 d.f.</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>North America</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Archaic Canada</td>
<td>65.2</td>
<td>60.0–71.1</td>
<td>56.6</td>
<td>50.0–67.3</td>
<td>0.0137</td>
<td>0.05–0.02</td>
</tr>
<tr>
<td>Iroquois Canada</td>
<td>56.6</td>
<td>50.0–67.3</td>
<td>52.4</td>
<td>45.0–59.8</td>
<td>0.0198</td>
<td>0.05–0.02</td>
</tr>
<tr>
<td>Maryland</td>
<td>41.8</td>
<td>31.1–52.5</td>
<td>52.4</td>
<td>45.0–59.8</td>
<td>0.0198</td>
<td>0.05–0.02</td>
</tr>
<tr>
<td>Athapascan</td>
<td>52.4</td>
<td>45.0–59.8</td>
<td>56.6</td>
<td>50.0–67.3</td>
<td>0.0137</td>
<td>0.05–0.02</td>
</tr>
<tr>
<td>California</td>
<td>60.0</td>
<td>50.0–70.1</td>
<td>56.6</td>
<td>50.0–67.3</td>
<td>0.0137</td>
<td>0.05–0.02</td>
</tr>
<tr>
<td>Grand Gulch, Utah</td>
<td>76.9</td>
<td>67.1–86.6</td>
<td>76.9</td>
<td>67.1–86.6</td>
<td>0.0104</td>
<td>0.05–0.02</td>
</tr>
<tr>
<td>Chelly and Kayenta</td>
<td>57.1</td>
<td>45.0–69.2</td>
<td>76.9</td>
<td>67.1–86.6</td>
<td>0.0104</td>
<td>0.05–0.02</td>
</tr>
<tr>
<td>Chevez Pass</td>
<td>61.5</td>
<td>50.0–72.8</td>
<td>76.9</td>
<td>67.1–86.6</td>
<td>0.0104</td>
<td>0.05–0.02</td>
</tr>
<tr>
<td>New Mexico</td>
<td>39.4</td>
<td>30.0–49.8</td>
<td>45.9</td>
<td>37.5–54.3</td>
<td>0.0160</td>
<td>0.05–0.02</td>
</tr>
<tr>
<td>Grasshopper</td>
<td>47.3</td>
<td>37.5–54.3</td>
<td>45.9</td>
<td>37.5–54.3</td>
<td>0.0160</td>
<td>0.05–0.02</td>
</tr>
<tr>
<td>Pt. Pines early</td>
<td>45.9</td>
<td>37.5–54.3</td>
<td>45.9</td>
<td>37.5–54.3</td>
<td>0.0160</td>
<td>0.05–0.02</td>
</tr>
<tr>
<td>Pt. Pines late</td>
<td>43.6</td>
<td>31.1–56.2</td>
<td>45.9</td>
<td>37.5–54.3</td>
<td>0.0160</td>
<td>0.05–0.02</td>
</tr>
<tr>
<td>Arkansas</td>
<td>55.7</td>
<td>45.0–66.4</td>
<td>55.7</td>
<td>45.0–66.4</td>
<td>0.0160</td>
<td>0.05–0.02</td>
</tr>
</tbody>
</table>

The dentition of American Indians 2017
Table 20.1 (continued)

<table>
<thead>
<tr>
<th>Trait, tooth</th>
<th>Break point</th>
<th>Shovel UL1</th>
<th>Uto-Aztec UP1</th>
<th>Enamel ext. UM1</th>
<th>One-root UP1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>2-7/0-7</td>
<td>1-0-1</td>
<td>1-3/0-3</td>
<td>1/1-3</td>
</tr>
<tr>
<td>Alabama</td>
<td></td>
<td>67.7</td>
<td>127</td>
<td>3.4</td>
<td>178</td>
</tr>
<tr>
<td>Coahuila, Mex.</td>
<td></td>
<td>61.5</td>
<td>26</td>
<td>7.9</td>
<td>38</td>
</tr>
<tr>
<td>Cuicuilco and Tehuacan</td>
<td></td>
<td>51.9</td>
<td>52</td>
<td>2.8</td>
<td>72</td>
</tr>
<tr>
<td>Tlatelolco</td>
<td></td>
<td>65.2</td>
<td>132</td>
<td>5.3</td>
<td>151</td>
</tr>
<tr>
<td>No.</td>
<td></td>
<td>683/1237</td>
<td>130/1604</td>
<td>143/2114</td>
<td>177/1637</td>
</tr>
</tbody>
</table>

Panama

<table>
<thead>
<tr>
<th>Trait, tooth</th>
<th>Break point</th>
<th>Shovel UL1</th>
<th>Uto-Aztec UP1</th>
<th>Enamel ext. UM1</th>
<th>One-root UP1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>61.9</td>
<td>63</td>
<td>11.4</td>
<td>70</td>
</tr>
</tbody>
</table>

Ayalan, Ecuador

<table>
<thead>
<tr>
<th>Trait, tooth</th>
<th>Break point</th>
<th>Shovel UL1</th>
<th>Uto-Aztec UP1</th>
<th>Enamel ext. UM1</th>
<th>One-root UP1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>66.2</td>
<td>68</td>
<td>7.8</td>
<td>77</td>
</tr>
</tbody>
</table>

Cotocollao, Ec.

<table>
<thead>
<tr>
<th>Trait, tooth</th>
<th>Break point</th>
<th>Shovel UL1</th>
<th>Uto-Aztec UP1</th>
<th>Enamel ext. UM1</th>
<th>One-root UP1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>52.2</td>
<td>23</td>
<td>12.5</td>
<td>40</td>
</tr>
</tbody>
</table>

Chanduy, Ec.

<table>
<thead>
<tr>
<th>Trait, tooth</th>
<th>Break point</th>
<th>Shovel UL1</th>
<th>Uto-Aztec UP1</th>
<th>Enamel ext. UM1</th>
<th>One-root UP1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>35.7</td>
<td>14</td>
<td>15.8</td>
<td>19</td>
</tr>
</tbody>
</table>

Santa Elena, Ecuador

<table>
<thead>
<tr>
<th>Trait, tooth</th>
<th>Break point</th>
<th>Shovel UL1</th>
<th>Uto-Aztec UP1</th>
<th>Enamel ext. UM1</th>
<th>One-root UP1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>80.0</td>
<td>10</td>
<td>5.7</td>
<td>35</td>
</tr>
</tbody>
</table>

Paloma, Peru

<table>
<thead>
<tr>
<th>Trait, tooth</th>
<th>Break point</th>
<th>Shovel UL1</th>
<th>Uto-Aztec UP1</th>
<th>Enamel ext. UM1</th>
<th>One-root UP1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>43.1</td>
<td>51</td>
<td>1.5</td>
<td>66</td>
</tr>
</tbody>
</table>

Peru 1 & 2

<table>
<thead>
<tr>
<th>Trait, tooth</th>
<th>Break point</th>
<th>Shovel UL1</th>
<th>Uto-Aztec UP1</th>
<th>Enamel ext. UM1</th>
<th>One-root UP1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>61.3</td>
<td>119</td>
<td>8.6</td>
<td>197</td>
</tr>
</tbody>
</table>

Chile

<table>
<thead>
<tr>
<th>Trait, tooth</th>
<th>Break point</th>
<th>Shovel UL1</th>
<th>Uto-Aztec UP1</th>
<th>Enamel ext. UM1</th>
<th>One-root UP1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>42.2</td>
<td>45</td>
<td>10.7</td>
<td>75</td>
</tr>
</tbody>
</table>

Corondo, Brz.

<table>
<thead>
<tr>
<th>Trait, tooth</th>
<th>Break point</th>
<th>Shovel UL1</th>
<th>Uto-Aztec UP1</th>
<th>Enamel ext. UM1</th>
<th>One-root UP1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>64.1</td>
<td>39</td>
<td>16.1</td>
<td>56</td>
</tr>
</tbody>
</table>

Lagoa Sta, Brz.

<table>
<thead>
<tr>
<th>Trait, tooth</th>
<th>Break point</th>
<th>Shovel UL1</th>
<th>Uto-Aztec UP1</th>
<th>Enamel ext. UM1</th>
<th>One-root UP1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>54.2</td>
<td>48</td>
<td>9.5</td>
<td>63</td>
</tr>
</tbody>
</table>

Sambaqui, Brz

<table>
<thead>
<tr>
<th>Trait, tooth</th>
<th>Break point</th>
<th>Shovel UL1</th>
<th>Uto-Aztec UP1</th>
<th>Enamel ext. UM1</th>
<th>One-root UP1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>60.9</td>
<td>69</td>
<td>6.2</td>
<td>113</td>
</tr>
</tbody>
</table>

No.

<table>
<thead>
<tr>
<th>Trait, tooth</th>
<th>Break point</th>
<th>Shovel UL1</th>
<th>Uto-Aztec UP1</th>
<th>Enamel ext. UM1</th>
<th>One-root UP1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>316/549</td>
<td>72/811</td>
<td>66/1100</td>
<td>61/796</td>
</tr>
</tbody>
</table>

North America

<table>
<thead>
<tr>
<th>Trait, tooth</th>
<th>Break point</th>
<th>Shovel UL1</th>
<th>Uto-Aztec UP1</th>
<th>Enamel ext. UM1</th>
<th>One-root UP1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td></td>
<td>55.2</td>
<td>8.1</td>
<td>6.8</td>
<td>10.8</td>
</tr>
<tr>
<td>Range</td>
<td></td>
<td>41.8–76.9</td>
<td>2.7–25.0</td>
<td>2.3–12.3</td>
<td>3.6–18.2</td>
</tr>
</tbody>
</table>

South America

<table>
<thead>
<tr>
<th>Trait, tooth</th>
<th>Break point</th>
<th>Shovel UL1</th>
<th>Uto-Aztec UP1</th>
<th>Enamel ext. UM1</th>
<th>One-root UP1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td></td>
<td>57.5</td>
<td>8.9</td>
<td>6.0</td>
<td>7.6</td>
</tr>
<tr>
<td>Range</td>
<td></td>
<td>43.1–80.0</td>
<td>1.5–16.1</td>
<td>2.8–16.8</td>
<td>0.0–19.5</td>
</tr>
<tr>
<td>NA/SA $\chi^2$ 1 d.f.</td>
<td>0.85</td>
<td>0.42</td>
<td>0.697</td>
<td>6.02</td>
<td></td>
</tr>
</tbody>
</table>

$^a$ Individual count, sexes pooled, all prehistoric except Athapaskan sample and some Iroquois. Sample provenience and break points are detailed in Turner (1985) and elsewhere. Scoring is done with ASU DAS (Turner et al. 1991).
Space limitations permit reviewing only a few of the 29 traits, so mainly those that distinguish Northeast Asian Sinodonty from Southeast Asian Sundadonty are presented (Turner 1983, 1987, 1990a). Figures 20.1 and 20.2 illustrate five of these traits—incisor shoveling, double-shoveling, tuberculum dentale, and first and second lower molar root number.

**Figure 20.1**
Arrow 1 indicates right upper central incisor shoveling trait in a female skull from prehistoric Alabama site Lu25–425. Arrow 2 points to double-shoveling. Not discussed in this report, but one of the key traits used in the multivariate analysis, arrow 3 points to canine tuberculum dentale (CGT neg. no. 6-6-80:19).

![Figure 20.1](image)

**Figure 20.2**
Arrow 1 indicates left lower first molar with two roots in prehistoric Peruvian male from a cemetery in the Chicama Valley. Had there been a supernumerary third root, it would be out of view on the distolinguinal aspect. Not discussed in this report, but one of the key traits, arrow 2 points to a one-rooted lower second molar (CGT neg. no.7-3-80:7. Reprinted courtesy of Dental Anthropology Newsletter).
20.5.1 Findings

American Indian dental morphological variation, assessed against the background of archeological and linguistic information, led to the following inferences:

1. There are three dental clusters of Macro-Indians: North America, South America, and an interregionally convergent group.
2. Dental variation is relatively low among Macro-Indians.
3. The Macro-Indian dental divisions arose by local evolution.
4. South America was colonized soon after North America.
5. Individual dental trait frequencies show only random variation.
6. Only Northeast Asian Sinodonty is present in the New World.
7. Dental variation supports the Clovis/epi-Clovis prehistory and Greenberg language migration models.

1. **Three dental clusters**: Figure 20.3 is a dendrogram computed with Ward’s clustering method for the 106 possible pairwise Mean Measures of Divergence, a multivariate similarity/dissimilarity statistic developed by C.A.B. Smith (Berry and Berry 1967) with widely used adjustments for sample size and determination of statistical significance (references in Turner 1985). There are three primary dental groups: North America (at the top), South America (at the bottom), and a middle group that shows interregional convergence.

   The Paloma sample is surely a misclassification attributable to chance. Some interregional convergence is expected given that dental microevolution (in the sense of gene frequency changes) is probably best explained as caused largely because of random changes due to founder’s effect, population structure, and local genetic drift. There is no archeological or other evidence to believe, for example, that the Ecuadorian Santa Elena series is similar to the Alabama sample because of ancestry closer than to each sample’s respective regional relationships. Instead, the Santa Elena-Alabama similarity is most likely an example of the occasional convergences that should be expected if Macro-Indian dental evolution was mainly due to chance, not selection or gene flow. As will be discussed below, the relatively low amount of variation, despite the two main North and South American clusters, is not supportive of multiple Paleo-Indian migrations.

2. **Low dental variation**: Inspection of Table 20.1 reveals a low amount of trait frequency variation between the 29 groups relative to that in, say, eastern Asia, or even New World Arctic people (Turner 1991). This limited New
World Indian dental divergence is probably what causes the rather limited treeness of the dendrogram in Figure 20.3. The limited treeness suggests either substantial gene flow throughout the Americas or a relatively recent colonization of the Americas with only minor in situ dental differentiation. Excellent treeness has been obtained throughout the New World (Arctic, Subarctic, non-Arctic) with these same dental traits and statistical methods (Turner 1987).

Table 20.1 shows that upper central incisor shoveling and lower first molar root number have very different overall frequencies in the New World. The former is a high frequency trait, the latter, a low frequency trait. Scott (1973) demonstrated that trait frequency was positively correlated with trait expression in the offspring of specific mating types, providing strong evidence for polygenic inheritance. As can be seen, shoveling is universally high (>70%) among the groups, whereas three-rooted lower first molars are uniformly at a low frequency, averaging about 6%. Both traits have little variation within and between North and South America. These two traits,
like the other 27, show no sign of clinal variation from north to south, east to west, coastal to interior, or low to high elevation. Trait frequencies have no identifiable gross environmental correlates in the New or Old World that would suggest the effects of short term natural selection.

3. Dental divisions arose by local evolution: Table 20.1 also lists the North and South American means and ranges for eight traits, seven of which distinguish Sinodonty from Sundadonty, plus a trait found almost exclusively in Native Americans—the Uto-Aztec premolar. The tabulation shows that there are no statistically significant differences in some traits, and very little in all others except the enamel extension, between pre-Columbian North and South American Indians. None of these traits has a north–south mean difference greater than 6% (enamel extensions). There is no tendency to exhibit less “Mongolization” or Sinodonty in South America than in North America, that is, there is no evidence for Sundadonty. The differences between the means appear to be mainly random. What little difference there is between North and South America is better interpreted as due to post-colonization local evolution rather than to predifferentiated multiple migrations. Moreover, the archeology, craniology, and odontology the senior author has personally seen and read about in Russian sources for eastern Siberia, does not provide any cultural or biological basis for hypothesizing markedly differentiated source populations in Primorye, Chukotka, trans-Baikalia, or Yakutia. However, the fact that these Siberian geographic districts and their pre-Russian cultures are recognizable today and prehistorically, could mean that they were also distinctive even earlier in Late Pleistocene times as well, and would have served as incubator habitats for some amount of pre-Beringian biocultural and linguistic differentiation.

4. No genetic bottlenecking at Panama: Because there are few dental differences between North and South American Indians, again, relative to the smaller area of eastern Asia, it would appear that there was no meaningful Paleo-Indian genetic bottlenecking in Panama. The size of the Paleo-Indian population wave that advanced through the isthmus was large enough to contain a representative sample of the North American Indian dental gene pool. A trait-by-trait comparison of presumed North and South American Paleo-Indians (Turner 1992a) turns up no differences greater than the within-continent range of trait occurrence. However, as is well known from various studies of blood group and DNA markers (Spuhler 1979; Szathmary 1979; Shields et al. 1992), the initial Paleo-Indian gene pool probably did not contain as representative a sample of Chukotka genes, let alone the Pan-Northeast Asian gene pool. For example, allele B of the ABO system was seemingly not carried across the Bering land bridge by Paleo-Indian
colonists, and the three-rooted lower first molar gene(s) was carried by only a small proportion of the first Beringian migrants.

5. **South America colonized soon after North America:** In 1976, MacNeish suggested that North America was populated for some 100,000 years, while South America had been inhabited for only 25,000 years. Had this been the actual occupational history of the New World, and if continuous occupation and regular postcolonization dental microevolution are assumed (Turner 1986), then the North American samples should have exhibited in Figure 20.3, on the basis of time alone, about fourfold more internal dental divergence than do the samples from South America. That is, the secondary branching in North America should be much further to the left compared with South America. The dendrogram provides no support for unequal evolutionary time in North and South America, nor do the individual trait values in Table 20.1. Admittedly, variation in the rate of dental microevolution, connected as it surely must be to population structure, size, and demography, does not allow precise estimates of separation between branching populations, in this case North and South America. However, claims for 25,000–30,000 years of South American occupation are certainly pushing the envelope of credibility on the basis of both New World and Old World dental evolutionary considerations and most archeological evidence (Lynch 1991; Haynes 2002a; Fiedel 2004). In the Old World it is a question of antiquity of Sinodonty, which I suspect is around whatever age is finally agreed upon for the Upper Cave teeth, so far ranging between 11,000 and 30,000 years BP (Chen et al. 1992).

6. **Individual dental trait frequencies show only random variation.** Comparisons of the means for North and South American dental trait frequencies (Table 20.1) show no identifiable trends that could be attributed to multiple migration, differential geographic selection, or some other type of localized natural selection directly affecting teeth. While these samples are not ideal for rigorously assessing the likelihood of selection pressure, the more obvious possible correlates are absent. Hence, the dental samples of the South American Pacific coastal populations of Chanduy-Valdivia, Santa Elena, Paloma, Peru, and Chile are not especially similar according to the analysis in Figure 20.3. These groups might be expected to be relatively similar due to some form of coastal environmental selection, but if this were so, then they should have also incorporated the coastal Sambaqui of Brazil.

Samples from upland environments are not any better correlated than the coastal groups. Mountain-plateau-originating Grasshopper and Point of Pines 1 are linked to coastal Peru. Mountain-located Cotocollao clusters with coastal-lowland Panama. High rainfall and tropical coast Sambaqui folk
are multivariately linked to low rainfall Arizona plateau Chavez Pass. The two long-lasting deep-winter Archaic Canada and Iroquois samples are joined, but their link with Coahuila from the cool winter high desert of northern Mexico does not suggest cold selection. Only city state organization versus noncity state seems to be a source of possible selection. Thus Tlatelolco, a late and large central Mexican Aztec metropolis, clusters with the similarly developed Peruvian samples. Their clustering together may be the result of oral and other disease selection associated with agricultural economies and highly processed cereal and tuber foods. However, if caries selection had been responsible, then city state-level Tlatelolco and Peru should have higher frequencies of simpler and potentially more caries-resistant teeth, such as 4-cusped lower second molars, less shoveling, and fewer first molar cusps, which is not the case for either sample. Still, disease selection needs further study with respect to dental morphology because caries are rare in hunters and gatherers but very common among agriculturalists.

7. Only Sinodonty is present in the New World.

a. Shoveling: ◁ Figure 20.4 illustrates upper central incisor shoveling. ◁ Tables 20.1 and ◀ 20.2 show crown and root trait frequencies within

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure20.4.png}
\caption{Circum-Pacific and European frequencies of upper central incisor shoveling}
\end{figure}
### Table 20.2

World dental frequency variation for the eight distinguishing Sinodont and Sundadont morphological traits

<table>
<thead>
<tr>
<th>Trait tooth Break point</th>
<th>Shovel UI1 2-7/0-7</th>
<th>Double-shovel UI1 2-6/0-6</th>
<th>1-root UP1 1/1-3</th>
<th>Enamel extension UM1 1-3/0-3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arctic (^b)</td>
<td>78</td>
<td>75</td>
<td>95</td>
<td>46</td>
</tr>
<tr>
<td>Eastern USA and Canada</td>
<td>91</td>
<td>78</td>
<td>80</td>
<td>31</td>
</tr>
<tr>
<td>SW USA</td>
<td>94</td>
<td>93</td>
<td>89</td>
<td>51</td>
</tr>
<tr>
<td>California</td>
<td>94</td>
<td>90</td>
<td>84</td>
<td>42</td>
</tr>
<tr>
<td>Mesoamerica</td>
<td>92</td>
<td>90</td>
<td>87</td>
<td>49</td>
</tr>
<tr>
<td>So. America</td>
<td>92</td>
<td>90</td>
<td>87</td>
<td>49</td>
</tr>
<tr>
<td>No. China-Mongolia</td>
<td>84</td>
<td>30</td>
<td>77</td>
<td>51</td>
</tr>
<tr>
<td>Recent Japan</td>
<td>66</td>
<td>20</td>
<td>75</td>
<td>55</td>
</tr>
<tr>
<td>Recent Thailand</td>
<td>37</td>
<td>9</td>
<td>66</td>
<td>39</td>
</tr>
<tr>
<td>Early Malay Archipelago</td>
<td>30</td>
<td>28</td>
<td>68</td>
<td>18</td>
</tr>
<tr>
<td>Melanesia</td>
<td>9</td>
<td>5</td>
<td>57</td>
<td>4</td>
</tr>
<tr>
<td>Australia</td>
<td>20</td>
<td>1</td>
<td>58</td>
<td>9</td>
</tr>
<tr>
<td>West Europe</td>
<td>2</td>
<td>4</td>
<td>58</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Trait tooth Break point</th>
<th>Peg-Reduced-C.A. UM3 prc/norm+prc</th>
<th>Deflecting wrinkle LM1 1-3/0-3</th>
<th>3-root LM1 3/1-3</th>
<th>1-root LM2 1/1-3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arctic (^b)</td>
<td>20</td>
<td>30</td>
<td>30</td>
<td>31</td>
</tr>
<tr>
<td>Eastern USA and Canada</td>
<td>18</td>
<td>45</td>
<td>6</td>
<td>32</td>
</tr>
<tr>
<td>SW USA</td>
<td>21</td>
<td>35</td>
<td>6</td>
<td>29</td>
</tr>
<tr>
<td>California</td>
<td>17</td>
<td>45</td>
<td>6</td>
<td>32</td>
</tr>
<tr>
<td>Mesoamerica</td>
<td>19</td>
<td>28</td>
<td>6</td>
<td>29</td>
</tr>
<tr>
<td>So. America</td>
<td>25</td>
<td>38</td>
<td>6</td>
<td>37</td>
</tr>
<tr>
<td>No. China-Mongolia</td>
<td>53</td>
<td>29</td>
<td>34</td>
<td>42</td>
</tr>
<tr>
<td>Recent Japan</td>
<td>42</td>
<td>35</td>
<td>24</td>
<td>33</td>
</tr>
<tr>
<td>Recent Thailand</td>
<td>18</td>
<td>19</td>
<td>11</td>
<td>31</td>
</tr>
<tr>
<td>Early Malay Archipelago</td>
<td>0</td>
<td>11</td>
<td>6</td>
<td>33</td>
</tr>
<tr>
<td>Melanesia</td>
<td>13</td>
<td>18</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Australia</td>
<td>5</td>
<td>23</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>West Europe</td>
<td>12</td>
<td>7</td>
<td>1</td>
<td>27</td>
</tr>
</tbody>
</table>

\(^a\)Individual count, sexes pooled, historic and prehistoric native groups, sample provenience, and break points detailed in Turner (1985) and elsewhere.


U denotes upper; L, lower; CA, congenital absence.
the Americas, eastern Asia, Melanesia, Australia, and western Europe. Shoveling is very common throughout the Americas and in Northeast Asia. There is no significant difference in the shoveling frequencies of North America (91.7%) and South America (91.9%). It is less common in Southeast Asia, Australia, much less so in Melanesia, and very rare in western Europe. This trait alone shows that Paleo-Indians more likely originated in the north China-Mongolia gene pool than in those of the other areas. That gene pool, in turn, had to have had its morphogenetic origin in the Sundadont dental pattern of Southeast Asia and South China—the closest dental pattern in the world to Sinodonty. Clearly, Paleo-Indians did not originate in Europe, Oceania, or Southeast Asia according to the distribution of the incisor shoveling genes. The African dental pattern is too different from that of Sinodonty to be considered relevant to Native American origins considerations (Turner 1992b).

The frequency of shoveling in the earliest North and South American crania is high (ca. 90.1%) (Turner 1992a p 18), nearly identical to that of recent populations (Table 20.1). This apparent similarity is not what would be expected for the stated implication that the Kennewick skeleton was not a Sinodont (Powell and Rose 1999), as well as the claim that early South American skeletons were also not Sinodonts (Lahr 1995 p 163).

b. Enamel extensions: Figure 20.5 shows the frequencies for the enamel extension polymorphism, a quasi-continuous trait found on the buccal surface of the upper first permanent molar. While shoveling might conceivably have some minor adaptive value (Mizoguchi 1985), it is difficult to imagine how selection could favor the tiny extension of enamel on the subgingival root surface. In fact, these smooth enamel extensions could have a slightly negative value because they do not provide a porous surface for periodontal tissue attachment, hence, favoring the formation of peridontal disease pockets in the adjacent alveolar bone. Enamel extension variation provides essentially the same frequency picture as shoveling; namely, extensions are common in Indians and Northeast Asian Sinodonts, slightly less so in Southeast Asian Sundadonts and very uncommon in Melanesians, Australians, and Europeans.

c. One-rooted upper first premolar: Figure 20.6 shows the frequencies for the upper first (P3 in paleontological notation) premolar root number polymorphism, which can have one to three roots. As with the shoveling and enamel extensions, one-rooted upper first premolars are common in Northeast Asian Sinodonts and all pre-Columbian Indians, less common
### Figure 20.5
Circum-Pacific and European frequencies of upper first molar enamel extension

<table>
<thead>
<tr>
<th>Region</th>
<th>Frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eastern US/Canada</td>
<td></td>
</tr>
<tr>
<td>Southwest US</td>
<td></td>
</tr>
<tr>
<td>California</td>
<td></td>
</tr>
<tr>
<td>Mesoamerica</td>
<td></td>
</tr>
<tr>
<td>South America</td>
<td></td>
</tr>
<tr>
<td>N. China/Mongolia</td>
<td></td>
</tr>
<tr>
<td>Recent Japan</td>
<td></td>
</tr>
<tr>
<td>Recent Thailand</td>
<td></td>
</tr>
<tr>
<td>Early Malay Archipelago</td>
<td></td>
</tr>
<tr>
<td>Melanesia</td>
<td></td>
</tr>
<tr>
<td>Western Europe</td>
<td></td>
</tr>
</tbody>
</table>

### Figure 20.6
Circum-Pacific and European frequencies of one-rooted upper first premolars (P3)

<table>
<thead>
<tr>
<th>Region</th>
<th>Frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eastern US/Canada</td>
<td></td>
</tr>
<tr>
<td>Southwest US</td>
<td></td>
</tr>
<tr>
<td>California</td>
<td></td>
</tr>
<tr>
<td>Mesoamerica</td>
<td></td>
</tr>
<tr>
<td>South America</td>
<td></td>
</tr>
<tr>
<td>N. China/Mongolia</td>
<td></td>
</tr>
<tr>
<td>Recent Japan</td>
<td></td>
</tr>
<tr>
<td>Recent Thailand</td>
<td></td>
</tr>
<tr>
<td>Early Malay Archipelago</td>
<td></td>
</tr>
<tr>
<td>Melanesia</td>
<td></td>
</tr>
<tr>
<td>Western Europe</td>
<td></td>
</tr>
</tbody>
</table>
in Southeast Asian Sundadonts, and slightly less common in Melanesians, Australians, and western Europeans. While there is less continental Old World and South Pacific islander occurrence of one-rooted upper first premolars, there is nothing in the data to suggest that American Indian ancestry was anything other than from Northeast Asian Sinodonts.

d. Deflecting wrinkle: Figure 20.7 shows the frequencies of another tiny secondary trait, the first permanent molar deflecting wrinkle. This poly-

![Figure 20.7: Circum-Pacific and European frequencies of lower first molar deflecting wrinkle](image)

morphism is the degree of distalward deflection from none to pronounced of the medial ridge of cusp 2 (mesiolingual cusp). This feature has almost no potential for adaptation as it is usually worn off the first molar by the beginning of reproductive age. The deflecting wrinkle is common in pre-Columbian American Indians and Northeast Asian Sinodonts. It is less frequent in Sundadonts and western Europeans. Melanesians and Australians are highly variable for the deflecting wrinkle, and the present Oceanic samples have a substantial frequency, fitting the geographic expectation of an old Southeast Asian origin for the ancestors of Pacific Islanders.
e. Three-rooted lower first molar: Figure 20.8 shows the frequencies for the three-rooted lower first permanent molar, another root polymorphism with one to three possible roots. Lower molar and upper premolar root number are statistically unrelated morphogenetic features. By far, worldwide, the two-rooted condition is most common for the lower first molar root number. The oldest known example of a three-rooted lower first molar in anatomically modern humans is the 22,000-year-old mandible fragment from the Tabon Caves site on Palawan Island in the western Philippines, excavated by Fox (1970). The fragment has three root sockets at the first molar location. Three-rooted lower first molars are a Sinodont characteristic, but, like the missing B allele in living American Indians, Paleo-Indian colonists almost failed to carry to the New World the gene(s) responsible for this accessory root on the lower first molars.

20.5.2 Dentition supports Ice-free Corridor, epi-Clovis/Clovis first, and Macro-Indian models

Unlike the other dental traits discussed here, three-rooted lower first molar frequency is less than that of Northeast Asian Sinodonts, and this trait is generally

![Figure 20.8](image)

Circum-Pacific and European frequencies of three-rooted lower first molars

<table>
<thead>
<tr>
<th>Region</th>
<th>Frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eastern US/Canada</td>
<td></td>
</tr>
<tr>
<td>Southwest US</td>
<td></td>
</tr>
<tr>
<td>California</td>
<td></td>
</tr>
<tr>
<td>Mesoamerica</td>
<td></td>
</tr>
<tr>
<td>South America</td>
<td></td>
</tr>
<tr>
<td>N. China/Mongolia</td>
<td></td>
</tr>
<tr>
<td>Recent Japan</td>
<td></td>
</tr>
<tr>
<td>Recent Thailand</td>
<td></td>
</tr>
<tr>
<td>Early Malay Archipelago</td>
<td></td>
</tr>
<tr>
<td>Melanesia</td>
<td></td>
</tr>
<tr>
<td>Western Europe</td>
<td></td>
</tr>
</tbody>
</table>
uncommon although uniformly present at about 5% in pre-Columbian Indians. These characteristics imply two events. First, the initial number of Siberians to reach eastern Beringia was apparently small and not strongly representative for all genetic characters. Second, the trait’s Panamerican uniformity suggests (1) that after crossing the ecologically patchy Bering land bridge, group size increased significantly because similar dental gene samples were carried southward to all other parts of the New World; and (2) all American Indians discussed here are descended from the founding Siberian epi-Clovis migrants whose archeological record for colonizing Alaska and Chukotka begins about 12,000 BP (Goebel et al. 2003). The demographic events of rapid population growth and widespread territorial expansion, leading to genetic stabilization in North and South America, could have occurred first in interior Alaska or later after subsequent generations reached and passed through the inhospitable and limiting corridor between the Cordilleran and Laurentide ice sheets of Canada. Unfortunately, there is nothing that we can identify in the dental data that helps us better understand where and how the rapid expansion began, regardless of how one defines migration (i.e., wavelike, pulsed, clonal spread, chaotic drift, niche-based, continuous leakage, leap-frogging, etc.), or which migration route one chooses (“ice-free corridor,” Pacific coast, ice-crossings and polar desert trek to Atlantic coast, or some combination of these). However, given the rarity of Paleo-Indian sites in Alaska, the recent Fairbanks and Brooks Range finds notwithstanding (West 1981, 1996; Powers et al. 1990; Hoffecker et al. 1993; Kunz and Reanier 1994; Yesner 2001; Yesner and Pearson 2002, others), we lean toward the demographic growth, related faunal extinctions (Martin 1990), and genetic drift stabilization as having explosively started at the southern exit of the western Canadian ice sheet corridor. We prefer the corridor entry over the coastal route because of the severe boating difficulty of getting past the Late Pleistocene glacial ice mass on and around the Alaska Peninsula (Elias 2002; Turner 2002; Hoffecker and Elias 2003). In contrast to the rarity of Clovis or Clovis-like fluted points along the entire Pacific coast of North America, Carlson (1991) identified at least 40 archeological sites in the ice-free corridor area of British Columbia and Alberta that had various types of fluted points.

Elsewhere the senior author (Turner 1992a) inventoried a number of dental observations from crania that were “candidates” for Paleo-Indian chronometric status. None of these incomplete individuals deviated from the Sinodont pattern. Since then, he has examined the teeth of four other crania that are dated as Paleo-Indian or Early Archaic. These include Sulphur Springs woman, excavated in southern Arizona by Waters (1986), Horn Rock Shelter double burial near Waco, Texas, and the Wilson-Leonard female, also from Texas (Young et al. 1987; Steele 1989; Steele and Powell 1992). These four also conform to the Sinodont dental
pattern as best as can be determined given the considerable amount of tooth wear they and other hunter–gatherers worldwide exhibit.

Finally, some archeologists (reviewed in Dixon 1999; for opposition see Carlson 1991; Haynes 2002b) and geneticists (discussed previously) have argued for a Pacific coastal entry route to North and South America from Siberia despite the absence of archeological evidence for, and Alaska Peninsula glacial evidence against, such a route (Workman 2001; Turner 2003). Moreover, Macro-Indian language family distributions in the Americas, Penutian, for example (Ruhlen 2000), do not suggest Pacific coastal entry. As for dental morphology, the Pacific coast samples presented in Table 20.1 show no sign of meaningful differences with interior samples as would be expected, had there been an earlier more Sundadont-like migration or a fourth American variant of Sinodonty.

**20.6 Concluding discussion**

In assaying the different types of biological data brought to bear on the peopling of the Americas, there are disagreements on the numbers of migrations and their timing, but there are many points of agreement as well. The homogeneity among American Indians indicated by dental morphology is paralleled by mtDNA and Y chromosome haplogroup data. The dentition shows a dichotomy between North American and South American Indians and this is also evident in genetic markers. Several studies critique the three migration model of Greenberg et al. (1986), but these often fail to include data on Eskimo-Aleut populations. If researchers argue that mtDNA and Y chromosome data support a position that Eskimo-Aleuts differentiated from American Indians after the arrival of a common ancestor in the New World, there are serious problems with their data sets (or the interpretation thereof). On the basis of teeth, nuclear markers, and even craniometry, Eskimo-Aleuts are consistently more closely aligned with recent Asian populations than are American Indians.

The placement of other northern groups, especially Na-Dene speakers, is a bit less certain than the Eskimo-Aleut case. Based on similarities in mtDNA, Shields et al. (1993) concluded that Eskimo-Aleuts and Athapaskans were very closely related, diverging from one another in the American Arctic as recently as 7,000 years ago. This same general position has long been advocated by Szathmary (1979, 1981, 1993; Szathmary and Ossenberg 1978) based on the variation in blood group and serum protein polymorphisms. Cavalli-Sforza et al. (1994) found Na-Dene groups fell between Eskimo-Aleuts and Amerinds in general. This intermediary position is also indicated by dental morphological variation (Turner 1985; Scott 1991; Scott and Turner 1997). When ancillary fields are taken
into account, especially linguistics and archeology, it is hard to reconcile common origins for Eskimo-Aleuts and Na-Dene speakers in the Holocene. Even rare genetic markers speak against this purported tie—Albumin Algonkin (formerly Albumin Naskapi) is found in polymorphic frequencies in Athapaskans and Algonquians but not in Eskimos (Lampl and Blumberg 1979). Eskimos have the B allele and Subarctic Indians do not (Harper 1980). The Y chromosome haplogroup C-M130 has also been found exclusively in Athapaskan and Algonquian populations (Schurr 2004a).

The level of genetic diversity in the Americas has led some authors to conclude that the New World was peopled long before the Last Glacial Maximum, with many estimates exceeding 30,000 years BP. To a considerable extent, this flies in the face of what is known about the archeology of Siberia and Beringia, let alone Australia whose aborigines’ biology, tools, and language have evolved so much that few resemblances remain with their Southeast Asian homeland. The New World founding population had to have been in place in northeast Asia before any groups could start budding off to colonize the Beringian landscape. In a recent synthesis of mtDNA and Y chromosome analyses, Schurr (2004a) proposes three major peopling events from Asia to the Americas: (1) the initial founding population in the Americas came from south-central Siberia and arrived in the New World between 20,000 and 14,000 cal year BP; given the presence of ice-sheets across the breadth of Canada, the route of this migration is presumed to be coastal; (2) a second migration, following an interior route, contributed to many of the populations of North and Central America; and (3) Beringian populations, including the ancestors of Aleuts, Eskimos, and Na-Dene speakers, came into New World following the Last Glacial Maximum.

Although Schurr’s reconstruction parallels in a number of ways our dental findings, we still contend that the degree of dental differentiation in the New World favors a late entry model, a position more in line with current archeological knowledge (Haynes 2002a; Fiedel 2004) and some genetic studies (cf. Seielstad et al. 2003). In principle, we have no objection to an earlier date for the peopling of the New World. At present, however, we do not feel a case has been made for this position, especially in light of archeological success in Australia in finding very early sites by a much smaller number of archeologists and geologists (Jelinek 1992). If a coastal migration did take place, more evidence for this event is required. For example, of all the pre-Clovis archeological sites listed by Schurr (2004a) to support early entry, none are along the Pacific coast. Scholars will continue to find “pre-Clovis” sites and develop molecular clocks and models of linguistic differentiation that indicate early human entry into the Americas, but the final arbiter of dispersal will come from archeological sites that have excellent stratigraphy and no dating issues.
The dental characteristics of pre-Columbian American Indians easily fit with the hypotheses of a rapidly expanding, Clovis- or epi-Clovis first colonization event, long advocated by Martin (1990) and Haynes (1991, n.d.), and the Macro-Indian language evolution model developed by Greenberg (1990). Because all prehistoric and unadmixed living Native Americans, including Na-Dene Greater Northwest Coast and Aleut-Eskimo, only briefly discussed here, have the Sinodont dental pattern, it would seem that when a date for the emergence of full-blown Sinodonty in Asia is established with some certainty, then that will have to be the earliest possible date for the subsequent colonization of the Americas, assuming that early and late similarity actually means genetic continuity. Inasmuch as the Upper Cave crania seem to have a Sinodont dental pattern, then whatever date is finally settled on for that assemblage will provide a reasonable time estimate for the potential colonization of the Americas.

We focus on Upper Cave since there are only a few sites in Siberia with Late Pleistocene human remains. One, near Lake Baikal, called Mal’ta, seems to have European- rather than Asian-like teeth (Turner 1990b). Two sites west of Lake Baikal have physical anthropological signs of Mongoloid or Sinodonty. These are the Late Pleistocene Yenisei River sites in and near Krasnoyarsk. In the city is Afontova Gora, the river bank section from which came a fragment of a subadult frontal bone that the late Russian physical anthropologist Alekseev (1998) believed to have been Mongoloid because of the size and form of the adhering nasal bones. Upstream ca. 35 km (21 mi) is Listvenka, from which came a mandible of a child whose unerupted first molar is slightly more Asian than European in overall appearance. Hence, broadly speaking, the pre-Arctic ancestral homeland of Paleo-Indians must have been in north China, Mongolia, and southern Siberia. It is easy to envision the newly evolved Sinodonts quickly expanding into northeastern Siberia, after they succeeded in domesticating the dog for hunting and hauling, perhaps drifting north out of China via the Vitim River system. Since there were no known resident Siberians above the Arctic Circle, the ancestral Paleo-Indian northward drift would have been rapid, and without human resistance in Beringia, and equally so southward from eastern Beringia. Because the dental differences between the north China-Mongolia group and all unadmixed Indians is so small compared with the much larger difference between Northeast and Southeast Asians, we can only conclude that the relatively small amount of intratrait dental variation within the New World reflects the simple evolutionary fact that Sinodonts have been in the Americas for a relatively short period of time, less time than it took for Sinodonty to evolve out of Sundadonty. Moreover, the colonists and their dogs were so reproductively successful that the usual genetic drift cause of short-term dental trait frequency change was reduced or negated by the large population size that quickly grew south of the east Beringian Arctic
steppe. This evolutionary scenario, despite years of bioarcheological research, does not differ much from that first proposed on craniological grounds by Hrdlička (1925). While some readers may find such lack of theoretical and empirical change as unthinkable in the rapidly changing world of science, others, ourselves included, recognize it as a tribute to Hrdlička’s empirical orientation, and one of the more probable scenarios in the complex world history of Late Pleistocene human microevolution and dispersal. For the present, Macro-Indian dental variation is not supportive of a Pacific coastal entry route to North and South America, leaving the late entry Ice-free Corridor model as dentally most parsimonious.

Acknowledgments

The dental data were collected with aid provided by the National Science Foundation, National Geographic Society, Arizona State University, IREX, US National Academy of Sciences, and the (former) USSR Academy of Sciences, including the Siberian Branch. Many of the data were collected with the help of Jacqueline A. Turner and Korri Dee Turner, with computer data processing by Linda S. Watson and Rhea Jacanin. Many individuals and institutions are represented, and each has been acknowledged in other publications. Figures 20.1 and 20.2 were photographed with permission of the Departments of Anthropology, respectively, of the US National Museum of Natural History (Douglas Ubelaker) and the University of Alabama (Kenneth Turner).

References


Carter GF (1957) Pleistocene man at San Diego. Johns Hopkins Press, Baltimore
Haeussler AMF (1996) Dental anthropology of Russia, Ukraine, Caucasus, Central Asia: the evaluation of five hypotheses for Paleo-Indian origins. Ph.D. dissertation, Department of Anthropology, Arizona State University, Tempe


Harris EF (1977) Anthropological and genetic aspects of the dental morphology of Solomon islanders, Melanesia. Ph.D. dissertation, Department of Anthropology, Arizona State University, Tempe


Haynes CV Jr (n.d.) Clovis progenitors: from Swan Point, Alaska to Anzick site, Montana in less than a decade? Manuscript in author’s possession


Hooton EA (1930) The Indians of Pecos Pueblo. Yale University Press, New Haven


Martin PS (1990) 40,000 years of extinctions on the “planet of doom.” Palaeoecogr Palaeoclimatol Palaeoecol 82: 187–201


Merriwether DA, Hall W, Vahle A, Ferrell RE (1996) mtDNA variation indicates Mongolia may have been the source for the founding population of the New World. Am J Hum Genet 59: 204–212


Nichol CR (1990) Dental genetics and biological relationships of the Pima Indians of Arizona. Ph.D. dissertation, Department of Anthropology, Arizona State University, Tempe


Schanfield MS (1992) Immunoglobulin allotypes (GM and KM) indicate multiple founding populations of Native Americans: evidence of at least four migrations to the New World. Hum Biol 64: 381–402


Starikovskaya YB, Sukernik RI, Schurr TG, Kogelnik AM, Wallace DC (1998) MtDNA diversity in Chukchi and Siberian Eskimos: implications for the genetic history of
Torroni A, Schurr GG, Yang C-C, Szathmary EJE, Williams RC, Weiss KM, Wallace DC (1992) Native American mitochondrial DNA analysis indicates that the Amerinds and the Nadene populations were founded by two independent migrations. Genetics 130: 153–162
Turner CG II (1990a) The major features of Sundadonty and Sinodonty, including suggestions about East Asian microevolution, population history, and late Pleistocene relationships with Australian Aboriginals. Am J Phys Anthropol 82: 295–317


Turner CG II, Bennett BH, Burnett SE (n.d.) The effect of wear on scoring of dental crown trait occurrence and expression (in preparation)


21 Overview of Paleolithic Archeology

Nicholas Toth · Kathy Schick

Abstract

The Paleolithic, or Old Stone Age, comprises over 99% of human technological history and spans a time range from 2.6 Ma (the earliest recognizable stone tools and archeological record) to 10,000 years ago (the end of the last ice age). There are three major stages of the Paleolithic: (1) The Early Paleolithic which includes: (a) The Oldowan, from 2.6 to about 1.0 Ma, characterized by simple core forms on cobbles and chunks (choppers, discoids, polyhedrons), battered percussors (hammerstones and spheroids), flakes and fragments, and retouched forms such as flake scrapers. Cut marks and fracture patterns on animal bones indicate meat and marrow processing, with the use of simple stone knives and hammers. This stage is associated with the later australopithecines and the earliest forms of the larger-brained genus Homo and documents the first hominid dispersal out of Africa and into Eurasia and (b) The Acheulean, which lasted from approximately 1.7 Ma to 250,000 years ago, and was characterized by large bifaces such as hand axes, cleavers, and picks. The early Acheulean is associated with Homo erectus/ergaster, while the later Acheulean (by ca. 500,000 years ago) is associated with the even larger-brained Homo heidelbergensis. (2) The Middle Paleolithic/Middle Stone Age, from about 250,000 to 30,000 years ago, characterized by a focus on retouched flake tools, such as scrapers, points, and backed knives, and prepared core technologies such as the Levallois method. The controlled production and use of fire appears to be widespread for the first time. This stage is especially associated with archaic forms of Homo sapiens (having modern-size brains but more robust faces and postcranial skeletons), including the Neandertals and the earliest anatomically modern humans. (3) The Late Paleolithic, from 40,000 until 10,000 years ago, characterized by blade tool industries, a proliferation of artifacts in bone, antler, and ivory, and the emergence of rich symbolic art in the form of paintings, engravings, sculpture, and personal body adornment. Early examples of clear architectural structures, musical instruments, and mechanical devices (spear-throwers and bows and arrows) emerge during this time.
This stage is especially associated with anatomically modern humans, *Homo sapiens sapiens*.

### 21.1 Introduction

The Paleolithic is the term applied to a very broad, early period of human prehistory beginning with the first archeological evidence of stone toolmaking approximately 2.6 Ma, through to the end of the Pleistocene epoch about 10,000 years ago, when the last continental glaciation receded. It is important to appreciate that over 99% of human technological development took place during the Paleolithic. The Paleolithic thus constitutes the bulk of the time span of human technological development and human prehistory and documents the emergence and evolution of the genus *Homo*. The term is applied primarily to prehistoric developments in the Old World, as the New World’s earliest archeological evidence appears only toward the very end of Paleolithic times, during the last phases of the terminal Pleistocene glaciation. In the New World, however, the period of Late Ice Age hunter-gatherers is often referred to as “Paleoindian,” and is contemporaneous with the last few thousand years of the Paleolithic in the Old World.

“Paleolithic” literally means the “Old Stone” (paleo = old, lithic = stone) Age, as it represents the earliest phases of human technological development when the vast majority of the tools represented in the archeological record were made of stone. At the end of the Pleistocene, the Paleolithic is followed by the later phases of the Stone Age, the Mesolithic and then the Neolithic. During the Mesolithic (in some regions referred to as the “Epipaleolithic”), stone technologies continued to evolve as stone tool-using hunter-gatherers adapted to changing environments of the current (Holocene) epoch, sometimes characterized by small (microlithic) stone tools. During the last phase of the Stone Age, often referred to as the Neolithic (or “New Stone” Age), a transition occurred from hunting-gathering to a more settled way of life based on food production (agriculture and herding), but stone continued for some time to be used for tools (such as ground axes, projectile points, and sickles).

The Paleolithic is traditionally divided into three major subdivisions: (1) the Early Paleolithic (also sometimes called the Lower Paleolithic) or Early Stone Age (ca. 2.6 Ma to 250,000 years ago; (2) The Middle Paleolithic or Middle Stone Age (ca. 250,000–30,000 years ago); and the Late Paleolithic (also Upper Paleolithic) or Later Stone Age (ca. 40,000–10,000 years ago). The “Lower”/“Middle”/ “Upper” designations for the Paleolithic stages were developed in Europe in the
late nineteenth and earlier twentieth centuries, based primarily on diagnostic artifact types and technological patterns observed in the stratigraphic and cultural sequences in various regions of Europe. More recently, with the appreciation that other parts of the world did not follow the precise cultural-historical sequence of Europe, many researchers have put less formal emphasis on these designations in favor of the more neutral terms “Early”/“Middle”/“Late” on a worldwide scale. This latter terminology will be used here.

For the first hundred years of Paleolithic research, these Paleolithic subdivisions were used to express a general chronological sequence (a relative chronology) without a firm sense of how many years ago each phase began or ended (an absolute chronology). During the past half-century, however, radiometric dating techniques have allowed the development of a more precise chronological framework for this Paleolithic sequence worldwide, with approximate times for the beginning and end of each phase.

Change from one stage of the Paleolithic to the next, however, does not always entail an immediate or complete turnover in artifact types, though it does generally represent an obvious and perceptible shift in the types of artifacts dominating the archeological tool assemblages, and often a corresponding shift in the dominant methods used in making these tools. For instance, while modified flake tools are present at a number of Lower Paleolithic sites, they become the dominant artifact form, often with consistent or repeated shapes, at many Middle Paleolithic sites. There is also some regional variation in the absolute chronology of the sequence, with evident technological transitions in some regions occurring earlier or later than in other regions. For instance, the transition from the Middle Paleolithic/Middle Stone Age to the Late Paleolithic/Later Stone Age happens somewhat earlier in some regions than in others.

21.2 Perspectives on early stone tools

The earliest prehistoric archeological record is now approximately 2.6 Myr old, based on the recognition of flaked stone artifacts in securely dated deposits in East Africa. The fossil record of bipedal hominids, however, goes back at least 6 Ma, several Myr before the first appearance of stone tools (see Volume 3 Chapter 5). On the basis of modern primate analogs, especially from chimpanzees, a range of tools and tool-using behaviors might be postulated for hominid populations prior to 2.6 Myr. Such hypothetical early tool use likely involved highly perishable, organic raw materials that provide no enduring, visible archeological record.
A handful of nonhuman species have been documented to show some minimal use of tools in the wild, including sea otters, birds (such as crows, finches, Egyptian vultures), and even mud wasps. Aside from humans, however, the only other animals showing habitual use of a variety of tools for a variety of purposes are our closest living relatives, the chimpanzees. What is more, chimpanzee toolmaking and tool-using skills appear to be learned over several years, suggesting a simple culturally transmitted system.

We now know that there is variability among different chimpanzee groups in the sets of tools they commonly use, showing cultural variation among chimpanzees in their tool kits. Modern chimpanzee tool use includes nut cracking with stone and wood hammers and anvils, termite fishing, and ant dipping with sticks or grass stems, and using chewed-up wads of leaves as sponges to obtain water or for self-cleaning. Although some chimpanzee tools consist of unmodified objects used for a particular task, chimpanzees do intentionally modify or shape some of their tools, such as the sticks and grasses used for termite fishing or ant dipping, and the chewed leaves used as sponges.

Deliberately manufactured stone artifacts in the early archeological record represent the earliest evidence of tool production by early hominids. As such, they reveal the development of a reliance on stone tool use in early hominid adaptation by at least 2.6 Ma. Although stone tool use may have been affected by seasonal, environmental, or other opportunities, the archeological record reveals a consistent manufacture of stone tools that persisted from this time onward until recent times.

Early stone artifacts clearly indicate a number of interesting behavioral characteristics of these early hominids: they selected stone raw materials at specific locations, transported manufactured artifacts and unmodified stone from one place to another on the paleolandscape, and discarded artifacts (and sometimes parts of animal carcasses) in distinct concentrations at many localities some distance from the raw material sources. Moreover, the manufacturing process used to produce early stone artifacts is one that is not observed in any nonhuman animal, even among chimpanzees, highlighting the novelty of behavioral innovation in the early stone toolmakers. Although early stone tools are admittedly simple and do not show elaborate shaping, they represent clear evidence of a new and unusual behavior pattern: the deliberate, controlled fracture of rock through percussive blows.

Technological patterns seen in early stone artifacts indicate they were produced primarily through a technique sometimes called “free hand, hard hammer percussion.” This involves hitting one rock (the hammer) against another (the core) to bring about controlled fracture of the core (called conchoidal fracture, as the shock waves can produce radiating, shell-like ripples in finer-grained
materials) and produce numbers of sharp pieces called flakes, a process called flaking or knapping. Experiments have shown that a main objective of early stone toolmaking was likely the production of such sharp flakes to use as cutting tools. Thus, a primary tool in the early hominid tool kit was likely the sharp-edged flake, and many of the cores found at early sites were likely by-products of the toolmaking process.

Early stone toolmaking hominids were consistently producing such fractured stones at a number of early site localities. Early Paleolithic sites often involve dozens of flaked cores and thousands of flake products. Analysis of early archeological materials often reveals extensive, controlled flaking of cores, involving rotation and manipulation to produce a series of flakes from the same piece of stone. Such fine core manipulation and exploitation is observable at even the very earliest Stone Age sites at Gona in Ethiopia, showing consistent, controlled, and skillful flaking of cores by 2.6 Ma.

With such skillful flaking observable among early hominid toolmakers on the one hand and the diverse tool-using and toolmaking cultures observable in chimpanzees on the other, a natural question is whether the production of early stone tools represents skills beyond those seen in other apes. Although chimpanzees are known to use stones as hammers and anvils in nut-cracking activities in West Africa, wild chimpanzee tool manufacture does not involve the intentional percussive flaking of stone, and wild chimpanzees have not developed sharp-edged tools for cutting in their assorted tool kits. It has been possible, however, to explore through experiments how comparable toolmaking skills of early hominids are to those of apes in captivity. An essential question in such experiments is whether the toolmaking skills of early hominids represent a significant departure from an ape “substrate” of toolmaking ability, and what insights we might gain regarding early hominid cognitive abilities. Do early hominid toolmakers exhibit special cognitive or biomechanical skills or abilities, or do these emerge only much later in human biological and technological evolution?

Experiments were begun in 1990 teaching a bonobo (pygmy chimpanzee), Kanzi, to make and use stone tools (Figure 21.1). The experiment involved introducing a use for a stone tool for cutting and retrieving a foodstuff, initial demonstration (modeling) of stone tool manufacture, and a subsequent period of trial-and-error learning on Kanzi’s part in both the toolmaking and tool-using operations. This experiment, which is still ongoing and now involves Kanzi’s younger sister and her two young offspring as well, has clearly shown that apes can become adept at some aspects of stone toolmaking skills and produce recognizable artifacts comparable to some of those found at early sites. Now, however, after more than 15 years of this ongoing experiment, some distinct technological differences persist in the bonobos’ artifacts compared to artifact
assemblages found at early Paleolithic sites. Some of these differences appear to reflect lesser skill in the bonobo toolmaker, perhaps reflecting lesser cognitive appreciation of particular facets of the toolmaking process (such as flaking sharper edges of the core, etc.), although others are likely related to biomechanical differences in the hand and arm of the apes.

This experiment highlights how skilled and adept early stone hominids were in their stone toolmaking by the time of the earliest known archeological occurrences 2.6 Ma. The skillfulness reflected in the earliest stone tools might...
indicate that even earlier stone technologies existed, yet undiscovered and perhaps rare on the paleolandscape, whose makers were not quite as proficient in flaking stone and who did not produce such a readily recognizable product. Or it may be that hominids were “preadapted” to efficiently flaking stone because of selection for other manipulative skills, which were later transferred to stone knapping when the need arose. The ape stone toolmaking experiments give important clues as to what technological characteristics might be found in such hypothetical “Pre-Oldowan” technologies.

### 21.3 Early Paleolithic

The Early Paleolithic comprises a long time interval, between 2.6 Ma and approximately 250,000 years ago. It not only includes this extremely large span of human prehistory but also encompasses, over time, sites across huge geographical distances, from southern Africa to eastern Asia. During this period of more than 2.25 Myr, profound evolutionary changes occurred among hominids, and some marked changes are observed in the archeological record in many parts of the Old World.

In Africa, where the Early Paleolithic is often referred to as the Early Stone Age, two industries have been recognized: (1) the first to appear, starting 2.6 Ma, the Oldowan Industry (named after Olduvai Gorge in Tanzania), consists of stone industries containing simple cores and flaked pieces, along with some battered artifacts such as hammerstones and (2) starting between 1.7 and 1.5 Ma, or approximately a Myr after the onset of Oldowan technology, the Acheulean Industry (named after the locality of St. Acheul in France) appears, with new, distinctive artifact forms in the form of relatively large bifacial tools (hand axes, cleavers, and picks).

#### 21.3.1 Oldowan

The Oldowan is the first recognizable archeological record, with simple flaked and battered stone artifacts, sometimes found with cut-marked and broken animal bones, emerging around 2.6 Ma. Although similar types of simple lithic industries are found throughout time, archeologists usually use a cut-off of around 1 Ma when referring to the Oldowan Industrial Complex. The Oldowan coexisted for several hundred thousand years with the Acheulean hand axe industries, starting about 1.7 Ma. Oldowan sites are known first from Africa, and subsequently document the spread of hominids outside of Africa into
parts of Eurasia, notably producing archeological sites in the Near East, the Republic of Georgia, and eastern Asia. These sites are found especially in tropical and subtropical climatic regimes, in particular grassland/woodland environments.

In East and North Africa, most Oldowan sites are open-air occurrences that are located along stream courses, in deltaic settings, or on lake margins. These were areas of close proximity to water and were depositional settings where sediments could build up over time. In South Africa, Oldowan artifacts are found in karstic limestone cave deposits and may have been carried there by hominids or brought in by natural forces such as slope wash or gravity. The high incidence of hominid bones in South African cave deposits (especially robust australopithecines) may be the result of predation and/or scavenging by carnivore such as leopards and hyenas.

Oldowan industries are contemporaneous with a number of bipedal hominid forms, including later australopithecines (Australopithecus garhi, A. aethiopicus, A. robustus, and A. boisei), whose cranial capacities ranged from about 400 to 550 cm³, and early forms of the more encephalized genus Homo (H. rudolfensis, H. habilis, H. ergaster/erectus), whose cranial capacities ranged from about 600 to 850 cm³. Although it is possible that all of these hominids used stone technology to a greater or lesser extent, many anthropologists believe that the genus Homo was probably a more habitual toolmaker and tool-user, as its brain size almost doubles in the first Myr of the Oldowan while its jaws and teeth tend to diminish in robusticity. By 1 Ma, only Homo ergaster/erectus was known in the human paleontological record, while the australopithecines became extinct. Interestingly, Homo ergaster/erectus appears to have much more modern limb proportions and stature relative to earlier hominids and is the first form clearly identified outside of Africa.

Oldowan industries are characterized by simple technologies (sometimes called Mode 1) consisting of cores made on pebbles or chunks (choppers, discoids, polyhedrons, heavy-duty scrapers, facetted spheroids), battered percussors (hammerstones and battered spheroids), debitage (flakes and fragments), and retouched pieces (scrapers, awls, etc.) (Figure 21.2). Common raw materials include volcanic lavas, quartz, and quartzite. The most common techniques for producing Oldowan artifacts were hard hammer percussion and bipolar technique (in which the core to be flaked is set on a stone anvil and hit with a stone hammer). At Olduvai Gorge, some technological trends have been observed through time, with later Oldowan sites showing higher frequencies of such artifact classes as scrapers and battered spheroids and lower frequencies of choppers. These sites are sometimes assigned to a “Developed Oldowan,” but such a designation is more difficult to apply elsewhere.
Microwear patterns on a small sample of Oldowan tools suggest that flakes were used for animal butchery, wood-working, and cutting soft plant matter. Experiments in using stone tools (Figure 21.3) have shown that Oldowan flakes can be used to efficiently process the carcasses of animals from the size of small mammals to elephants (Figure 21.4), and stone hammers could easily break bones for access to nutritious marrow and skulls for brain tissue. Choppers could have been used to chop branches to make spears or digging sticks, although many such Oldowan core forms were probably by-products of flake production. It is likely that a rich range of perishable organic material culture was also used, including containers of shell, horn, skin, or bark; wooden clubs and or throwing sticks; wooden spears or digging sticks (Figure 21.5); and horn or bone fragments as digging tools. In addition, a small sample of bone specimens from South African caves are polished and striated on their pointed end, suggesting that these may have been used as opportunistic digging tools to gain access to underground vegetable resources or insects such as termites.

Although evidence of fire has been found at a few Oldowan sites (in the form of reddened, baked sediments, burnt bones, or fire-cracked stone), it cannot be ruled out that natural agents, such as lightning strikes and brushfires, may have produced these fires. No clear architectural structures have been found at
Oldowan sites, and it is possible that Oldowan hominids could have been sleeping in trees at night (perhaps building nests like chimpanzees) rather than on the ground in order to avoid predation by nocturnal carnivores.

It seems clear that these Oldowan hominids were concentrating lithic material and animal bones at favored locations on the landscape (a pattern not seen in nonhuman primates today), but the precise behavioral patterns that formed these
concentrations are still debated. Interpretations for these concentrations include
home bases or central foraging places, favored places due to proximity to shade,
water, or food resources, intentional stone caches, and scavenged carnivore
accumulations. It is also possible that Oldowan sites formed through more than
one behavioral pattern. Cut marks and percussion marks/fractures on bones show
that hominids were accessing meat and marrow resources from animal carcasses
obtained through scavenging or hunting. The modified bones at Oldowan sites
typically come from animals ranging in size from small mammals to those

Figure 21.4
Butchery of an elephant, the world’s largest terrestrial mammal, using simple Oldowan
flakes. (The elephant had died of natural causes)

Figure 21.5
Sharpening a wooden branch with a simple stone flake. Such implements could have been
used as spears, digging sticks, or skewers to carry meat resources
weighing hundreds of pounds. This is a scale of carnivory that is not seen in the nonhuman primate world and was most likely greatly facilitated through the use of stone tools.

At present, there is a debate as to whether hominids accessed larger animals through more marginal scavenging (getting the ravaged leftovers of carnivore kills) or rather had access to more complete carcasses through hunting or confrontational scavenging. In any case, the processing of larger animal carcasses could have significantly increased the diet breadth (and thus survivorship and reproductive success) of Oldowan hominids, although the majority of Oldowan hominid diet was likely derived from plant foods such as fruits, berries, nuts, edible leaves, and underground storage organs (roots, tubers, corms, and rhizomes). Carrying devices may have facilitated the collection and transport of dietary items that could be consumed at a later time.

Important Oldowan localities include Gona, Fejej, and the Omo valley in Ethiopia; East and West Turkana in Kenya; Olduvai Gorge in Tanzania; Sterkfontein and Swartkrans caves in South Africa; Ain Hanech and el-Kherba in Algeria; the lowest levels at Ubeidiya in Israel; and Dmanisi in the Republic of Georgia.

### 21.3.2 Acheulean

The Acheulean Industrial Complex is characterized by the presence of large bifacial hand axes and cleavers (sometimes called Mode 2 technologies), which are found from approximately 1.7 Ma to 250,000 years ago. These hand axe/cleaver industries are contemporaneous and sometimes regionally co-occurring with the simpler Oldowan-like (Mode 1) industries. Acheulean and contemporaneous Mode 1 industries are found throughout Africa and Eurasia, but classic hand axe and cleaver assemblages are especially characteristic of Africa, the Near East, the Indian subcontinent, and western Europe. Elsewhere, notably eastern Europe and most of eastern Asia, simpler Mode 1, Oldowan-like technologies are found. This was a period of major climatic change, with numerous cold/warm oscillations that would have especially affected northern latitudes of Eurasia. For most of this period, hominids would have flourished only during the warmer periods in these northern latitudes. Hominids extended their range from grasslands and woodlands of tropical and subtropical regions to cooler, more temperate climates during this period.

Contemporaneous hominid forms include *Homo ergaster/erectus* and the later, larger-brained *Homo heidelbergensis* (sometimes referred to as “early archaic *Homo sapiens*). Cranial capacities range from about 800 to 1,400 cm³, generally increasing over the time span of this period. In the early Acheulean, robust
australopithecines (*A. robustus* and *A. boisei*) still existed, but most anthropologists do not regard these forms as plausible Acheulean toolmakers, and in any case they appear to have gone extinct by 1 Ma.

New elements in Acheulean industries (in addition to Mode 1, Oldowan-like artifacts that continue to be found) include hand axes, cleavers, picks, and knives (generically called “bifaces”) made either on large flakes struck from boulder cores or on larger cobbles and nodules. A range of well-made retouched tools, such as side scrapers, awls, and backed knives, are also common. Frequently used raw materials include fine-grained lavas, quartzites, and flints. Earlier, cruder bifaces were produced by hard hammer percussion (Figure 21.6), while later more refined bifaces were probably finished by the soft hammer technique, in which a softer material, such as wood, bone, ivory, antler, or even soft stone, was used as a percussor, producing thinner, more invasive flakes (Figure 21.7). Prepared core techniques, notably the Levallois tortoise core technique (in which a large, predetermined flake is removed from the upper surface of a discoidal core), and, more rarely, early blade production, are found in some later Acheulean industries. Sharpened wooden spears are known from later Acheulean times, as at Schöningen in Germany and Clacton in England, suggesting that
more formal hunting weaponry was established as part of a regular subsistence pattern by at least this time if not earlier.

The fact that Acheulean and contemporaneous hominids successfully occupied cooler, more temperate latitudes suggests that they were better adapted to such cooler conditions. Use-wear patterns on side scrapers indicate that many of these tools were used to scrape hides, strongly suggesting that animal skins were being used for simple clothing, blankets, and/or tent or hut coverings. Evidence of fire in the form of charcoal or ash layers is occasionally seen in later Acheulean times but is by no means widespread in the archeological record during this period. There is no definitive evidence of architectural structures during Acheulean times, although arguments have been made in this regard. Sites are found in numerous caves and rockshelters as well as many open-air sites.

Hand axes and cleavers in particular, indicate the ability to impose bilateral symmetry on lithic materials. This clearly shows higher cognitive abilities and motor skills than are manifested in the Oldowan. Even modern humans who learn to make stone tools normally require considerable apprenticeship before they can produce well-made hand axes and cleavers. Although there is a wide range of hand axe forms through time and space, it is common that at certain Acheulean sites there are recurrent shapes and sizes, as if there were stylistic norms of production among their makers. Presence of ochre at some sites and, occasionally, incised bone may indicate the emergence of proto-symbolic behavior as well.

Important Acheulean sites/localities include Konso-Gardula, Middle Awash, Melka Kunturé, and Gadeb in Ethiopia; Olduvai Gorge and Peninj in Tanzania; Olorgesailie and Isenya in Kenya; Kalambo Falls in Zambia; Elandsfontein and Montagu Cave in South Africa; Ternifine in Algeria; Ubeidiya and Gesher Benot Ya’aqov in Israel; Swanscombe, Hoxne, and Boxgrove in England; St. Acheul and Terra Amata in France; and Torralba and Ambrona in Spain. Important contemporaneous Mode 1 localities include Atapuerca (TD6) in Spain, Arago Cave in France, Clacton in England, Bilzingsleben and Schöningen in Germany, Vértesszöllös in Hungary, Isernia in Italy, and the Nihewan Basin and Zhoukoudian (“Peking Man”) cave in China.

21.4 Middle Paleolithic/Middle Stone Age

The Middle Paleolithic industries of Europe, the Near East, and North Africa (sometimes called the “Mousterian” after the site of Le Moustier in France) and Middle Stone Age industries of sub-Saharan Africa are found between approximately 250,000 and 30,000 years ago. They are found in tropical, subtropical, temperate, and even periglacial climatic regimes. During this time, hominids extended their ranges to most environmental zones of Africa and Eurasia except harsh deserts, the densest tropical forests, and extreme northern or arctic tundras. It appears that hominids were somehow able to cross the water between Southeast Asia and Australia, then attached to New Guinea and Tasmania, by late in this period. Contemporary hominid forms include those often designated as archaic Homo sapiens (including the Neandertals of Europe and the Near East) and anatomically modern humans.

Hand axes and cleavers tend to be less common (although toward the end of the Middle Paleolithic of Western Europe, smaller, well-made hand axes are found), and the emphasis of these stone industries is on retouched forms made on flakes (such as side scrapers, denticulates, and points) become numerous in
many of these assemblages (Figure 21.8). Hard hammer and soft hammer techniques were common during this period. Many of these industries exhibit prepared core methods, notably the Levallois technique for more controlled production of flakes, points, and sometimes blades. Wooden spear technology continues from the Acheulean (as seen at Lehringen, Germany, where a spear with a fire-hardened tip was associated with an elephant carcass), and stone points with tangs or thinned bases suggest that these forms may have been hafted onto spear shafts, suggesting the development of composite tools. Rare bone points are also known from this time.
Fire and hearth structures are much more common during this period, although clear architectural features outlined by stones or bones are rare. Sites are numerous in caves and rockshelters, as well as open-air sites on plateaus and along river floodplains.

Occasional perforated and grooved shells and teeth at a few sites imply the emergence of some personal adornment, and along with the infrequent presence of ochre or, more rarely, engraved bone as well as a number of well-documented burials, suggest at least some symbolic component to hominid behavior during this period of the Paleolithic.

Important Middle Paleolithic/Middle Stone Age sites include Combe Grenal, Pech de L’Azé, Le Moustier, La Quina, and La Ferrassie in France; Krapina in Croatia; Cueva Morin in Spain; Tabun, Skhul, Kebara, Amud, and Qafzeh in Israel; Shanidar in Iraq; Dar es Soltan in Morocco; Bir el Ater in Algeria; Haoua Fteah in Libya; Kharga Oasis in Egypt; Diré-Dawa, Omo-Kibish, and Middle Awash in Ethiopia; Enkapune Ya Muto, Prospect Farm, and Kapthurin in Kenya; Kalambo Falls, and Twin Rivers in Zambia; and Florisbad, Border Cave, Klasies River Mouth Cave and Die Kelders Cave in South Africa.

21.5 Late Paleolithic

The Late Paleolithic (often called Upper Paleolithic in Europe and Later Stone Age in Africa) is found between 40,000 and 10,000 years ago, at which time the last glaciation receded. This period of human prehistory overlaps and is contemporaneous with the end of the Middle Paleolithic/Middle Stone Age in some regions. During this time, humans inhabited tropical, subtropical, temperate, desert, and arctic climates; occupied present-day Australia, New Guinea, and Tasmania after crossing significant bodies of water; and, late in this period, spread to the Americas via the Bering Straits. Late Paleolithic industries are almost always associated with anatomically modern humans (Homo sapiens sapiens), but some early Upper Paleolithic sites in Europe are also contemporaneous with the last populations of Neandertals there.

Late Paleolithic stone industries are often characterized by blade technologies, elongated flakes produced by soft hammer or indirect percussion, in which a punch is placed on the edge of a blade core and struck with a percussor. These blades were then made into a variety of tool forms, including end scrapers, burins, and backed knives (Figure 21.9). Some Late Paleolithic technologies emphasized bifacial points, such as the Solutrean of Spain and France, and the Paleoindian occurrences of the New World (Clovis and Folsom). Such points may have been produced by soft hammer technique or by pressure flaking, in
which small flakes are detached by directed pressure rather than by percussion. Some raw materials appear to have been heat-treated to make them easier to work. Other Late Paleolithic technologies emphasized bladelets (small blades) and geometric microliths, which were hafted as composite tools into a range of projectiles and cutting tools. These microlithic technologies are characteristic of the Later Stone Age of Africa as well as some parts of central and eastern Asia.

A diagnostic element of many Late Paleolithic industries is an emphasis on nonlithic materials for tools, including bone, antler, and ivory, made into a range of artifact forms such as points, needles, spear-throwers, shaft straighteners, and harpoons. Hooked spear-throwers are essentially mechanical devices to increase the velocity and/or distance of a projectile, and thus represent a significant advance in hunting technology or weaponry. The small size of some points and
microliths toward the end of the Late Paleolithic suggest the development of bow and arrow technology, and arrows are preserved at Stellmoor, Germany.

Several human sculptures from the Late Paleolithic suggest clothing such as hooded parkas, headdresses, and aprons. The development of bone and antler needles also suggests that sewed clothing was common after 20,000 years ago, and recently discovered impressions on fired clay fragments from the Czech Republic indicate woven textiles, presumably of plant material.

Controlled use of fire appears to be a universal trait during this period, with hearths sometimes lined with stones. Architectural features are much more common than in earlier periods, with hut structures delineated by stone or bone patterns, by postholes, and sometimes with hearth structures and other apparent activity areas within (such as toolmaking or tool-using). Sites tend to be more numerous and have denser concentrations of materials, suggesting larger populations and more regular habitation of sites.

One of the most distinctive characteristics of the Late Paleolithic is the proliferation of symbolic expression in art and personal adornment (Figure 21.10). This can be seen in the naturalistic representation of animals and, more rarely, humans in painting and sculpture as well as in the more abstract geometric designs. A variety of media was employed for artistic expression, including use of charcoal, pigment paints, antler, bone and ivory, and clay, as well as a diversity of techniques, including drawing, painting, engraving, carving, and modeling. Personal adornments are sometimes numerous, manifested in beads or pendants of shell, bone, tooth, antler, ivory, and stone. This proliferation of symbolic expression, best seen in the European Upper Paleolithic, has sometimes been referred to as the “Creative Explosion.” Some of these artistic manifestations, particularly paintings, drawings, and engravings, are located in deep, hard-to-access recesses of caves, suggesting a ritualistic and religious aspect to this symbolism. In view of the complexity of the material culture of this period and its well developed symbolic component, it is likely that modern human language abilities were fully developed by this time, if not before.

Late Paleolithic burials are more common and more elaborate than in the Middle Paleolithic. Men, women, and children were sometimes interred with rich grave goods, including stone tools, jewelry, and bone/antler/ivory artifacts. Again, this suggests an important symbolic component and a probable belief in an afterlife, in other words, something akin to a spiritual belief and a religion.

Important sites include Lascaux, Pincevent, La Madeleine, Abri Pataud, Cro-Magnon, Solutré, Chauvet and Laugerie Haute in France; El Castillo, Altamira, and Parpalló in Spain; Dolní Vestonice in the Czech Republic; Vogelherd in Germany; Istállóskö in Hungary; Willendorf in Austria; Kebara Cave in Israel;
Ksar ‘Akil in Lebanon; Kostienki and Sungir in Russia; Mezin and Mezhirich in Ukraine; Mal’ta in Siberia; Zhoukoudian Upper Cave in China; Lukenya Hill in Kenya; Mumba Cave in Tanzania; Nelson Bay Cave, Die Kelders, Elands Bay Cave, and Wilton, in South Africa; Haua Fteah in Libya; Lake Mungo in Australia; and Blackwater Draw in New Mexico (North America).

Figure 21.10
Examples of probable symbolic behavior in Late Paleolithic times, expressed in art, personal adornment, music, notation, burial, and possibly more formal architecture

Early symbolism

- Venus figurine
- Bone flute
- Ivory beads
- Cave art: painting of reindeer
- Cave art: A possible shaman with flute
- Art and possible tally from antler fragment (unrolled)
- Burial
- Mammoth-bone hut

Overview of Paleolithic archeology
References


22 The Network of Brain, Body, Language, and Culture

Steven Mithen

Abstract
The evolution of the brain, language, body, and human culture are all intimately linked to each other and cannot be understood in isolation. While paleoanthropologists primarily consider the data from the fossil and archeological records, they must draw on recent findings and current research in a host of disciplines, including psychology, neuroscience, and human genetics. In this chapter, I review how the evolution of the brain, body, language, and material culture of humans have been linked during the course of human evolution, exploring how the interrelationships have changed during different periods of time. To do so, I divide this chapter into three main sections. First, I examine the Plio‐Pleistocene and stress the links between bipedalism, brain size, diet, and sociality. Second, for the Middle Pleistocene my emphasis shifts to a focus on the nature of protolanguage and the evolution of human life history. Finally, I deal with the origin and dispersal of modern humans for whom material culture begins to play a key role in the network between brain and behavior.

22.1 Introduction
The evolutionary relationship between the brain, body, language, and culture lies at the heart of paleoanthropology. Any interpretation of the archeological and fossil record ultimately relies on either implicit or explicit assumptions about the cognitive and communicative capacities of our ancestors and how these compare to our own today. Not only paleoanthropologists but also psychologists, neuroscientists, philosophers, and scholars from effectively any discipline that deals in some manner with human behavior have many different views as to how the brain, body, language, and culture are related. In this chapter, I will review and discuss a selection of these views while promoting particular scenarios for cognitive evolution and how this relates to body, brain, and behavior, extended
versions of which can be found in Mithen (1996, 2005). The scenario is one that I believe serves to explain the patterning and variability in the archeological and fossil records in the most parsimonious manner, while being consistent with our understanding from related disciplines such as developmental psychology and the philosophy of mind. As it will be focused on theoretical and interpretative issues, readers are recommended to consult the relevant chapters elsewhere in this three-volume handbook for details about the empirical data to which it refers.

22.2 A brief overview

The challenge facing paleoanthropologists is to explain how in approximately 6 Myr a primate probably not dissimilar to the chimpanzee evolved into *Homo sapiens*, existing today with a culture that includes the pursuit of pure mathematics, fundamentalist religious beliefs, medical science, and the internet. The paradox we face is that *H. sapiens* is genetically almost identical to the chimpanzee, our closest living relative with whom we shared that 6-Myr-old ancestor, but cognitively and behaviorally radically different.

Within that short period of time the genus *Homo* appeared, diverged into several evolutionary branches, colonized Eurasia, and then became restricted to the single species that survives today. More than 100 kyr after *H. sapiens* had evolved in Africa, it colonized Eurasia, and then the Americas in the Late Pleistocene and the Pacific Islands and Arctic regions in the Early Holocene. The first unambiguous traces of art, symbolic thinking, and complex technology appeared less than 100 ka, while a mere 10 ka all members of *H. sapiens* were still living as hunter-gatherers. It is, indeed, only since the appearance of farming in the Early Holocene that the full cognitive and behavioral potential of *H. sapiens* appear to have been realized, although the future may hold further surprises.

That potential must have its roots not only in the distinctive qualities of *H. sapiens* as a species but also in *Homo* as a genus. Hence, we require the whole of the paleoanthropological record to understand the nature of thought and behavior today. Unraveling the relationship between brain, body, language, and culture must lie at the center of that understanding and of any explanation for how an apelike primate evolved into the peculiar species that we are today.

Although the fossil record becomes richer year by year and occasionally throws up a surprise, such as the discovery of *H. floresiensis* in 2003 (Brown et al. 2004), the basic outline of human evolution is unlikely to be substantially altered by new discoveries, especially as it is increasingly supported by evolutionary genetics (Jobling et al. 2004).
In essence, there were numerous species of bipedal primates in Africa between 6 and 2 Ma. These display considerable morphological variation, which most likely relates to the exploitation of specific niches within the African landscape. These hominins have been placed into three genera, *Ardipithecus*, *Australopithecus*, and *Homo*, with the latter constituted by two purported species, *H. habilis* and *H. rudolfensis*. Flaked stone tools are known from at least 2.5 Ma, but in light of the repertoire of tools used by chimpanzees it seems likely that hammer stones, sticks, leaves, and other minimally modified materials were used long before flaked stone artifacts appeared.

The two early species of *Homo* are characterized by relatively larger brains than the other hominins, up to 650 cm$^3$ rather than the 450 cm$^3$, which is also characteristic of chimpanzees today, together with smaller teeth and a flatter face. These features may be within the range of variation for this grade of hominin without necessarily indicating any evolved cognitive or linguistic abilities. The key problem we face with assessing the significance of brain size is the rarity of the postcranial skeletons for earliest *Homo*, which might indicate that the relatively large brains are simply a product of large body size. Indeed, a strong case can be made for reclassifying *H. habilis* and *H. rudolfensis* as australopithecines and identifying *H. ergaster*, appearing by 1.8 Ma, as the first member of the genus *Homo* (Wood and Collard 1999).

*H. ergaster* may mark an evolutionary transition to a type of hominin for which behavioral analogies with living primates are of limited value. With a nearly fully modern stature and bipedal gait, this species is most likely the first to have dispersed out of Africa. Brain size reached up to 900 cm$^3$, although some specimens show the maintenance of relatively small brain capacities—those from the site of Dmanisi in Georgia were no more than 650 cm$^3$ (Gabunia et al. 2000; Lordkipanidze et al. 2000). The Asian lineage of this species evolved into *H. erectus* and appears to have made at least one water crossing on rafts to reach Flores Island by 850 ka (Morwood et al. 1998).

In Europe, the Middle Pleistocene is marked by a succession of hominins that are claimed to constitute at least three species: *H. antecessor*, *H. heidelbergensis*, and *H. neanderthalensis*. These appear to show a gradual increase of brain size until a capacity equivalent to, and in some cases exceeding, that of *H. sapiens* is attained in Late Pleistocene specimens. Stone artifacts appear to increase in technological complexity from Oldowan-like flakes associated with *H. antecessor*, Acheulian handaxes with *H. heidelbergensis*, and Levallois technology with *H. neanderthalensis*. Examples of bone and wooden tools are exceedingly rare, but the discovery of the Schöningen spears (Thieme 1997) indicates that this is most likely a consequence of preservation and discovery. Traces of structures and nonutilitarian artifacts are effectively absent, with the few ambiguous examples
claimed by some merely emphasizing their extreme rarity and unsophisticated nature.

The hominin species of Europe may form a single evolving lineage or multiple dispersals into that continent of species which evolved from *H. ergaster* in Africa. Within Africa there is evolutionary continuity from *H. ergaster* to *H. sapiens* (McBrearty and Brooks 2000) with the earliest specimens of the latter dating to c. 200 ka (McDougall et al. 2005), a date that effectively coincides with some estimates for this species appearance from the study of modern day genetic diversity (Ingman et al. 2000; Jobling et al. 2004). While *H. sapiens* specimens in Israel indicate initial dispersal out of Africa prior to 100 ka (Lahr and Foley 1994), the genetic evidence indicates that it was after 60 ka that major dispersals into Asia and Europe occurred and that it was these that gave rise to the extant populations today (Ingman et al. 2000).

Immediately prior to such dispersals we find evidence in South Africa for new types of material culture, often assumed to reflect the appearance of symbolic thought and language. Most notable are the incised ochre nodules and shell beads from Blombos Cave dating to 70 ka (Henshilwood et al. 2002, 2004), while red ochre is prevalent in Middle Stone Age deposits at South African sites reaching back to 100 ka (Knight et al. 1995).

The dispersals of *H. sapiens* out of Africa resulted in the colonization of Australia by at least 30 ka and possibly by 59 ka and that of Europe by 40 ka. The latter is associated with major technological innovations characterized by the Aurignacian culture that appears to arise in the Near East. The Neanderthals may have attempted to imitate the culture of the incoming *H. sapiens* (D’Errico et al. 1998), but they were either outcompeted for resources or unable to survive the major climatic fluctuations of the Late Pleistocene (D’Errico and Sanchez Goni 2003). By 30 ka, the *H. sapiens* in Europe were engaging in cave painting, carving intricate bone figurines, and making elaborately decorated burials. Similarly, the *H. sapiens* in Australia and Africa were likely to be engaged in both abstract and figurative rock art at this date.

The Neanderthals became extinct soon after 30 ka; precisely when *H. erectus* in Asia became extinct remains unclear; a similar date is likely, although *H. floresiensis* survived on Flores island until a mere 13 ka (Morwood et al. 2004). The climatic deterioration of the last glacial maximum of 20 ka caused *H. sapiens* to abandon northern landscapes that became polar desert and those areas in low latitudes that became extremely arid. As global warming began, *H. sapiens* recolonized those landscapes and proceeded to colonize the rest of the world either in the Late Pleistocene or Early Holocene—the far north, the Americas, and the islands of the Pacific (Mithen 2003).
It was only in the Holocene, beginning a mere 10 ka, that agricultural economies developed, initially in the Near East at c. 9 ka and then quite independently at several locations elsewhere in the world, including rice farming in China at 7 ka, maize and squash in Central America at 8,000 BC and domesticated camelids (llamas and alpacas) in the Peruvian Andes by 7,000 BC (Smith 1995; Mithen 2003). Farming provided the economic foundation for the development of towns and within a few thousand years the first “civilizations,” within which writing was independently invented. Since these developments, cultural complexity, economic efficiency, and population size appear to have grown exponentially, along with artistic achievement and scientific knowledge. The latter had grown so much by the nineteenth century that people began to examine their evolutionary past, which has led us today to enquire about the evolutionary relationships between brain, body, language, and culture.

To explore those relationships and how they have changed during the course of human evolution, I will divide this chapter into three sections. First, we will consider the australopithecines and earliest Homo, those hominins up to the emergence of H. ergaster at c. 1.8 Ma. The key issues in this section concern how these hominins adapted to environmental change and the network of evolving relationships between bipedalism, diet, sociality, toolmaking, and brain size. The second section will examine the large brained, fully bipedal hominins of the species H. ergaster, H. erectus, H. heidelbergensis, and H. neanderthalensis. Some anthropologists would want to add further species to this grade of hominin that I refer to as “Early Humans,” such as H. antecessor (Bermudez de Castro et al. 1997), while others would want to reduce their number by means such as considering H. ergaster and H. erectus to be a single species. The key issues to address are the evolving relationships between encephalization, vocal communication, intelligence, and dispersal into Eurasia. The final section will deal with H. sapiens and will contend that a fundamental change in the network between brain, body, language, and culture occurred with the evolution of this species.

22.3 The network of brain, body, language, and culture in the Plio-Pleistocene

22.3.1 Bipedalism, brain, and language

The first suite of relationships to examine is that relating to the evolution of bipedalism. How bipedalism originated has been debated for many years, with theories being gradually revised or dramatically falsified as new evidence has
appeared (Lewin 1999). It was once argued that bipedalism arose to “free the hands” so that stone artifacts could be manufactured. This had to be rejected when the evidence for australopithecine bipedalism was pushed back far beyond 2.5 Ma, the date at which the first stone tools were made. Similarly, the idea that bipedalism originated when hominins began living in savannah environments, perhaps to allow them to stand and watch for predators or prey over the top of long grass, could not cope with the evidence that the partially bipedal *Ardipithecus* and australopithecines inhabited wooded landscapes. The notion that bipedalism arose because male hominins needed to carry foodstuffs to the females with whom they had pair-bonded had to be similarly rejected when anthropologists recognized that the extent of body-size differences between males and females—the sexual dimorphism—indicated that such pair-bonding, and hence provisioning, is unlikely (McHenry 1996; Mithen 2005).

The most persuasive current argument for the origin of bipedalism is that it involved two distinct stages each with its own selective pressure, the first leading to the different bipedalism of the australopithecines and the second to the fully modern bipedalism of *H. ergaster*.

With regard to the first, Hunt (1994) argued that fruit consumption may have played a crucial role. He made a detailed study of the situations in which chimpanzees stand, and occasionally walk on two legs, and found that they do so principally to harvest fruit from small trees. Using their hands for balance, they pick fruit and slowly shuffle from tree to tree. By standing on two legs both hands can be used, while the bipedal shuffling avoids the need to keep raising and then lowering their body weight, which delivers a major saving in energy expenditure. Hunt suspects that the australopithecines were behaving in a similar fashion, arguing that Lucy’s anatomy was more suited to standing than walking. The broad pelvis, he suggests, made a stable platform, while the curved fingers and powerful arms should be seen as adaptations not for climbing trees but for hanging by one hand while standing on two legs and feeding with the other.

The “stage two” shift to a fully modern bipedalism is most likely related to the spread of savannah like environments in East Africa soon after 2 Ma owing to major climatic change (DeMenocal 1995). But rather than being driven by the need to see across tall grass or for carrying, it probably relates to the need to reduce heat stress owing to increased exposure to the sun in landscapes that had markedly fewer trees (Wheeler 1984, 1991, 1994).

Heat stress is a problem to all animals living in savannah environments, largely because brains begin to malfunction if raised 2°C above their normal temperature. By standing erect, a hominin would have only absorbed sunlight on the top of its head and shoulders; by knuckle walking its whole back would have been exposed. Moreover, the air is cooler away from the soil itself, and wind velocities
are significantly higher which would have improved the cooling effectiveness of bodily evaporation.

By using a range of computer and engineering models, Wheeler demonstrated that by walking bipedally, savannah living hominins would have significantly reduced both heat stress and water consumption. This may have played a major role in “opening-up” the new dietary niche of meat eating as such reductions allowed hominins to travel greater distances across savannah landscapes in search of carcasses, possibly at times of the day when competing carnivores would be resting in the sun.

Aiello (1996) built on Wheeler’s ideas to consider the relationship between bipedalism, brain, and language. She argued that brain expansion could be explained by the new demands on sensory-motor control that bipedalism both required and enabled. Standing or walking on two legs requires that the center of gravity is constantly monitored and small groups of muscles frequently recruited and changed to correct the body’s position; the movement of the legs has to be integrated with that of the arms, hands, and trunk to maintain a dynamic balance. And once those arms and hands are freed from a locomotory role, they can be used independently from the legs such as for carrying, signaling, throwing, and making tools. Bipedalism requires, therefore, a larger brain and more complex nervous system just to attain the more complex sensory-motor control. Once evolved for these reasons, the larger brain might then be used for other tasks including the planning of foraging behavior, social interaction, and language: intelligence may have been no more than a spin-off from walking on two legs.

The evolution of complex vocalization would have been additionally facilitated by the impact of a bipedal anatomy on the vocal tract. The human larynx is found much lower in the throat than that of the chimpanzees, and this enables a much wider array of sounds to be produced. Anthropologists had traditionally assumed that strong selective pressures for spoken language had “pushed” the larynx into the throat, and had the ill effect of creating a risk of choking on food. But Aiello (1996) argued that the low larynx was a consequence of the anatomical adaptations necessary for bipedalism rather than having been selected itself for enhanced vocalization. Because the spinal cord now had to enter the brain case from below rather than behind (as shown by the position of the foramen magnum), there was less space between the spinal cord and the mouth for the larynx. This space had been further reduced by the changes in the hominin face and dentition that arose with greater degrees of meat eating. Consequently, the larynx had to become positioned lower in the throat, which had the incidental effect of lengthening the vocal tract and increasing the diversity of possible sounds.
Aiello (1996) also suggested that not only the position but also the makeup of the larynx itself may have been changed by the new bipedal anatomy, resulting in the production of a less harsh, more melodious sound than in the australopithecines. The larynx of modern apes and humans is essentially a valve—the vocal cords—that can be closed off to allow the air to build up, and then opened to create a sound. This “valvular larynx” also has a quite different function, one to do with movement. Air pressure behind a closed larynx stabilizes the chest and provides a fixed basis for arm muscles—this is why we hold our breath when about to use our arms vigorously such as when throwing a ball. Chest stabilization is really important for primates who use their arms for knuckle walking or climbing. As a consequence, they have relatively thick, cartilaginous vocal cords that produce rather harsh sounding vocalizations. The vocal cords of modern humans, however, are more membranous.

Bipedalism may have caused the change by relaxing the selective pressure on the locomotor function of the valvular larynx. If so, then by becoming less rigid H. ergaster’s vocal cords may have further enhanced the diversity of sounds now being made by this bipedal hominin, even though there had been no selective pressures for enhanced vocalization itself.

22.3.2 Diet, body, and brain

The hominin body, brain, and vocal abilities were also involved in a network of evolutionary developments related to changes in diet, as explained in Aiello and Wheeler’s (1995) “expensive tissues” hypothesis. This was based on two key observations: first, that the basal metabolic rate of modern human is essentially the same as in nonhuman primates, and indeed all mammals scaled for body size; second, that owing to its size, the modern human brain has a much higher metabolic cost that of other mammals. The only manner in which these two observations can be reconciled is by modern humans having reduced metabolic costs in other aspects of their physiology.

Aiello and Wheeler examined a range of potential areas in which this might be found such as the functioning of the liver, heart, and lungs. But the metabolic rates of these organs are dependent on body size and hence could not have undergone a reduction in their metabolic rate to compensate for increased brain size. The only organ by which this could have been achieved is the gut, and this could only have been attained by an enhanced quality of diet. Aiello and Wheeler argued that by consuming high protein foods, such as meat, eggs, and fruit, the size and hence metabolic cost of the gut could have been reduced, which could then have compensated for the increased costs of the brain to maintain a
constant basal metabolic rate. Aiello and Wheeler found that in modern humans, the gut is indeed relatively reduced in size as compared to that of other nonhuman primates, reducing its metabolic cost by the appropriate amount to account for the high costs of the large brain.

The evolutionary time period of this change appears to have been the Pliopleistocene. The postcranial anatomy of the australopithecines—admittedly represented by a very small sample—suggests an extended gut as found in modern-day apes, while that of the *H. ergaster* specimen WT 15000, the Nariokotome Boy, indicates a gut comparable in size and bodily proportion to modern humans. This period also suggests a major change in diet from archeological rather than fossil evidence: the remains of stone artifacts and fragmented animal bones, both of which indicate increased meat consumption (Isaac 1978, 1983, 1989). Whether such meat derived from hunting or scavenging remains unclear (Binford 1986; Bunn and Kroll 1986), and is probably of limited significance, as both required planning and cooperation to pursue, while stone artifacts with sharp edges were required to butcher carcasses.

The shift to a greater meat component in the diet is evidently related to the evolution of bipedalism and brain size. As noted above, bipedalism enabled hominins to exploit a scavenging/hunting niche unavailable to other carnivores by allowing them extended foraging time. Bipedalism also enabled hominins to cover a larger foraging range and “freed the hands” for the manufacture and manipulation of stone artifacts and for carrying carcass parts. Larger brains—and hence computing power—were also essential. Hunting, scavenging, and tool manufacture required planning, transmission of information, and hand–eye coordination.

We have, therefore, a complex network of interrelated evolutionary changes. Increased meat-eating enabled a reduction in gut size; this enabled an increased brain size that provided the cognitive capacities to formulate more complex foraging strategies and to manufacture stone artifacts which enabled further increase in the consumption of meat and other high quality foods such as fruit. And it may have been the picking for fruit that provided the initial selective pressures for bipedalism, which was then further selected by the exploitation of an available scavenging niche and which, in turn, freed the hands for making stone tools and carrying.

### 22.3.3 Society, body, and brain

A further factor to enter into this network of causes and consequences in human evolution is sociality: the demands of living in larger social groups may have
imposed stronger selective pressures for an enlargement of the brain and enhanced communication than those arising from foraging and tool manufacture. Three lines of evidence suggest that group size increased during the Plio-Pleistocene: archeological remains, comparative ecology, and brain size.

From 2.5 Ma, archeological sites appear with extensive scatters of stone artifacts and fragmented animal bones, which were interpreted by Isaac (1978) as “home bases” (Potts 1988). Interpretation of such sites is extraordinarily difficult as when large quantities of artifacts are present, it is rarely possible to distinguish whether these were left by a large group of hominins or by multiple occupations of the same locality by a relatively small number of individuals. Moreover, there may have been large aggregations of hominins at localities and times when there was no requirement to manufacture stone artifacts or to butcher animal carcasses. Such social aggregations would leave no archeological trace and hence to claim that these only began when we have an archeological record appears dubious.

It is not, however, only the quantity of material at sites such as FLK 22, HAB and FxJJ50 (Isaac 1978; Bunn et al. 1980; Bunn and Kroll 1986) that suggests relatively large groups of hominins. As Isaac observed, the stone artifacts often imply that they had been transported from other parts of the landscape while the diversity of animal species suggest that they had originated from a range of environments from which carcass parts had been transported to a central location, most likely for food sharing. Isaac envisaged such sites as localities where small groups of hominins who spent much of the day apart would aggregate for food sharing, the care of young, and transmission of information.

Such aggregations may have also been a risk-minimizing strategy in face of potential predators, notably terrestrial carnivores. Comparative socioecology indicates that primate group size is higher in open than wooded environments (Dunbar 1988); without trees to escape into, a large group can deter predators and reduce the chance that any one individual will be preyed upon. Hence the increasing aridity from around 2 Ma, which led to a reduction in tree cover, is also likely to have led to increased hominin group size. The fossil record provides clear evidence that hominins were indeed preyed upon by carnivores (Brain 1981).

A third line of evidence comes from brain size. Dunbar (1993; Aiello and Dunbar 1993) has shown that there is a correlation between group size and brain size in nonhuman primates today. He argues that the increase in brain size between the australopithecines and Homo, and then within the Homo genus itself, reflects increasing group size. This is most likely correct for the earlier phases of human evolution, such as for H. habilis and H. ergaster, but when dealing with much larger brained hominins, such as H. neanderthalensis and H. sapiens—whose brain sizes are far outside the range of nonhuman primates—one cannot be confident that the brain size–group size correlation still applies. Nevertheless,
for the Plio-Pleistocene the increase in brain size appears to support the indications from archeological evidence and comparative socioecology that the earliest *Homo* were living in relatively larger groups than their australopithecine ancestors.

The implications of this for cognitive and cultural evolution are substantial. Whereas group living provides advantages for predator defense, food sharing, and information transmission, it also provides increased potential for competition and conflict (Dunbar 1988). To thrive within a large group one needs to monitor the behavior and relationships of a relatively large number of other individuals. This has led anthropologists to argue that there would have been selective pressures for the so-called “theory of mind” ability in Plio-Pleistocene hominins. This is a “mind-reading” ability that enables one to appreciate that another individual has beliefs and desires that differ to one’s own and hence modify one’s behavior accordingly. As such, theory of mind is central to the social lives of modern humans.

All nonhuman primates may lack mind-reading abilities and hence be unable to appreciate that another individual has beliefs and desires different to one’s own. Following the publication of Cheney and Seyfarth’s (1990) *How Monkeys See the World*, there is widespread agreement among academics that monkeys lack a theory of mind, but there is still considerable disagreement regarding chimpanzees (Povinelli 1993, 1999; Tomasello et al. 2003). Some believe that occasional observations of deceptive behavior suggest that chimpanzees can appreciate the difference between their own thoughts and those of another because deceiving involves providing another with false beliefs.

Byrne and Whiten (1988, 1992) championed this view, bringing together a great many isolated observations of apparently deceptive behavior to produce a strong case that chimpanzees have a theory of mind. As many of their observations were “anecdotal” reports, questions remained regarding the validity of this conclusion, and some (predominately laboratory-based) psychologists refused to accept it. But recent laboratory experiments have now demonstrated that chimpanzees can indeed understand psychological states, vindicating the inferences that Byrne and Whiten made from observations of “natural” behavior (Tomasello et al. 2003).

This certainly does not mean that the chimpanzee theory of mind is the same as that of humans. The term “theory of mind” must encompass a range of cognitive skills but is open to various interpretations (Carruthers and Smith 1996). There is, for instance, the ability to appreciate that another individual has a particular desire, such as for food, which may be different from appreciating that an individual has a particular belief. Another distinction concerns what has become known as the “orders of intentionality.” If I know what I think, then
I am termed as having a single order of intentionality; if I know what someone else thinks then I have two orders of intentionality; if I know what someone else thinks that a third partly thinks, then I have a third order of intentionality, and so forth. Whereas humans routinely use three or four orders of intentionality in their social life, apes might be limited to two orders at most (Dunbar 2004). Hence, although apes may have a theory of mind, it is of a quite different nature to that which humans possess. As a consequence they have a relatively limited understanding of what another individual is thinking, which must be a key factor underlying their very limited call repertoire.

The relatively large brains of *H. habilis* and *H. rudolfensis* when compared to the australopithecines and modern-day apes might reflect an enhanced theory of mind capability (Dunbar 2004). The selective pressure for this leap in understanding would have arisen from living in larger groups and the consequent increase in the complexity of social interaction (Mithen 1999). Those individuals who were more able to predict the behavior of others because they could “read their minds” would have had a competitive advantage within the group.

While the selective pressure may have primarily arisen from social life, the implications for toolmaking, foraging, and communication would have been profound. Although chimpanzees utilize a diverse range of tools, they appear constrained in the ease with which their methods of manufacture and use are transmitted from one generation to the next. Although the use of hammerstones, anvils, and termite sticks provide substantial nutritional advantages, juveniles take a considerable amount of trial and error to learn how to use them, receiving limited, if any, instruction from their parents and other experienced individuals (Boesch 1991). This may be because those older individuals cannot appreciate that juveniles lack the necessary knowledge about how to manufacture and use such tools.

With an enhanced theory of mind capacity the acquisition and transmission of technological knowledge may have been facilitated allowing the development of a stone tool technology. The manufacture of Oldowan tools dating from 2 Ma is considerably more complex than the production of termite sticks, ant-probes, or any of the other artifacts utilized by chimpanzees (Toth 1985; Mithen 1996). Social rather than trial and error learning is most likely essential, with individuals needing to appreciate the intentions in the minds of those whom they watched striking one stone with another. Similarly, with the possession of a theory of mind capacity, experienced stone knappers would have appreciated that juveniles lacked the necessary knowledge and hence may have been able to engage in active teaching.

The possession of a theory of mind may have also had considerable implications for the evolution of complex vocalizations and body gestures that laid the foundations for language. While the vocalizations of monkeys and apes may often
be intentional and referential (Zuberbühler 2003), the extent of information transmission appears largely constrained to that about emotional states. If individuals are unable to appreciate that the beliefs and desires of another individual are different from one’s own, then there will be limited need to communicate. Hence, although vervet monkeys have a repertoire of predator alarms calls, they appear to lack vocalizations relating to other activities that would appear to be of ecological value, such as from mothers to infants to indicate that they are moving on—infants sometime get left behind (Cheney and Seyfarth 1990). Conversely, if one appreciates that another individual has quite different beliefs and desires to one’s own, then there may be the need to communicate one’s own beliefs and desires either by vocalization, gesture, or body language in general. This may not only be about the social world—who are friends, allies, or mating partners with whom, but also about the natural world such as where game or carcasses might be located and about toolmaking. We would expect, therefore, that the social groups of early hominins were not only larger than those of their australopithecine ancestors but also rather noisier, with a greater diversity of vocalizations about a greater range of subjects (Mithen 2005).

22.3.4 Summary: hominin evolution in the Plio-Pleistocene

There was a complex network of evolving relationships between anatomy, brain size, foraging behavior, sociality, vocalization, technology, and environment during the Plio-Pleistocene. While a prime mover behind these developments was the changing environmental context of hominins, notably increasing aridity and the transition from a woodland to a savannah environment, it is difficult to distinguish between cause and consequence with regard to the behavioral and anatomical changes. Indeed, the most reasonable interpretation of the evidence is that there was feedback between each of these developments within a continuous nexus of evolutionary change that culminated in the appearance of *H. ergaster* at c. 1.8 Ma, with a lifestyle that resulted in the dispersal out of Africa.

22.4 The network of brain, body, language, and culture in early humans: *Homo ergaster, erectus, heidelbergensis, and neanderthalensis*

The chronological pattern of encephalization appears to show a period of increase between 2 and 1.5 Ma, and then more than a million years of stasis until a more
dramatic increase of brain size between 0.6 and 0.25 Ma (Ruff et al. 1997). That period of stasis occurred throughout fluctuating Pleistocene environments, and after hominins had dispersed into Asia and Europe. It suggests that they had achieved a very adaptable lifestyle based around the use of stone artifacts, hunting in the higher latitudes and large social groups. What brought this period of stasis to an end remains unclear to anthropologists. But the increase in brain size after 0.6 Ma is most likely related to the evolution of language.

22.4.1 The nature of protolanguage

The cultural accomplishments of early humans were substantial: dispersal out of Africa into both northern latitudes and Southeast Asia, production of complex stone artifacts such as handaxes and Levallois points, big game hunting, and survival during the dramatic environmental changes of the Middle Pleistocene. To have attained such accomplishments, some form of vocal communication more complex than that found in the great apes today appears essential.

Language is a particularly complex system of communication. Many believe that it has to have evolved gradually in a succession of ever more complex communication systems used by the early humans and their early Homo and australopithecine ancestors (Pinker and Bloom 1995). Academics refer to these communication systems by the catchall term “protolanguage” and identifying the nature of protolanguage is the most important task facing anyone attempting to understand how language evolved. Current theories for protolanguage fall into two “camps”—those who believe that protolanguage was “compositional” in character and those who believe it was “holistic.”

The essence of compositional theories is that protolanguage consisted of words with limited, if any, grammar. The champion of this view is Bickerton (1990, 1998, 2000) who argues that human ancestors and relatives such as the Neanderthals may have had a relatively large lexicon of words, each of which related to a mental concept such as “meat,” “fire,” “hunt,” and so forth. They were able to string such words together but could only do so in a near arbitrary fashion. Bickerton recognizes that this could result in some ambiguity. For instance, would “man killed bear” have meant that a man has killed a bear or that a bear has killed a man? Jackendoff (1999, 2000) suggests that simple rules such as “agent first” (i.e., the man killed the bear) might have reduced potential ambiguity. Nevertheless, the number and complexity of potential utterances would have been severely limited. The transformation of such protolanguage into language required the evolution of grammar—rules that define the order in
which a finite number of words can be strung together to create an infinite number of utterances, each with a specific meaning.

Compositional theories of protolanguage have dominated studies of language evolution for the past decade; they have been highly influential but have now been challenged by “holistic” theories, as proposed by Wray (1998, 2000, 2002) and Arbib (2002, 2003, in press). By using the term holistic, Wray means that the precursor to language was a communication system composed of “messages” rather than words; each hominin utterance was uniquely associated with an arbitrary meaning, as are the words of language today and indeed those of a Bickertonian protolanguage. But in Wray’s protolanguage, hominin multisyllable utterances were not composed of smaller units of meaning (i.e., words) that could be combined together in either an arbitrary fashion or by using rules to produce emergent meanings. In this view, modern language only evolved when such holistic utterances were “segmented” to produce words that could then be composed together to create statements with novel meanings. Arbib (2002) uses the term “fractionation” in a similar manner. Hence while Bickerton believes that words were present in the early stages of language evolution, Wray and Arbib believe that they only appeared in its latter stages.

Whether early human protolanguage was compositional, holistic, or some combination of both in character, the anatomical changes evident from the fossil record suggest that strong selective pressures had existed for enhanced vocal communication. The Kebara burial provided anthropologists with the first discovery of a Neanderthal hyoid bone, in fact the first from any premodern human (Arensburg et al. 1989, 1990). The hyoid bone is attached to the cartilage in the larynx and anchors the muscles necessary for speech. The Kebara specimen is morphologically similar that of a modern human and has been used to argue that the Neanderthal vocal tract was also essentially the same as that which you or I possess (see discussion between Arensburg et al. 1990; Lieberman 1990, 1992). It was certainly not identical and contrasts in facial morphology probably led to nasal-like vocalizations. The claim by Lieberman and Crelin (1971) that the Neanderthal larynx had been positioned high in the throat—equivalent to that of a modern human baby or a chimpanzee—have now been thoroughly dismissed. So although the Neanderthal vocal tract may have been unable to produce exactly the same range of sounds as that of a modern human, they would certainly have been sufficiently diverse to enable speech if the neural circuitry for language had been present within the Neanderthal brain.

Two further aspects of anatomical evidence support the claim for vocal capabilities equivalent to those of modern humans. First is the dimension of the hypoglossal canal, which carries the nerves from the brain to the tongue
(Kay et al. 1998). The width of this is larger in modern humans than in chimpanzees and gorillas, reflecting the larger supply of nerves required for the complex motor control necessary for speech. The canal is partially preserved at the base of the cranium; Kay et al. (1998) found that those from *Australopithecus africanus* and *H. habilis* specimens fell into the chimpanzee and gorilla size range, while those from Neanderthals were equivalent to the canal size of modern humans.

A second nerve-carrying canal measurable from fossilized remains is that of the thoracic vertebrae through which the nerves for controlling the diaphragm, and hence breathing, pass. MacLarnon and Hewitt (1999) found that this has had a similar evolutionary history to the hypoglossal canal: the vertebral canals of australopithecines and *H. ergaster* are similar in size to those of African apes, while those from Neanderthals are larger and match the size of those from *H. sapiens*. MacLarnon and Hewitt considered and rejected explanations for its increase in size relating to bipedalism, breathing for running, or the avoidance of choking on food. Their conclusion was that the increased number of nerves passing through the thoracic vertebrae arose for an increased control of breathing for speech.

The evidence from the hypoglossal and vertebral canals indicates that Neanderthals had a similar motor control of their tongues and breathing as modern humans. This also appears to be the case for the perception of sound, owing to evidence from the inner ear. While australopithecines appear to have been more sensitive to higher frequency sounds than are modern humans (Moggi-Cecchi and Collard 2002), the sound perception of *H. heidelbergensis* equivalent to modern humans appears to have been equivalent to that of modern humans (Martínez et al. 2004). As *H. heidelbergensis* was most likely a direct ancestor of the Neanderthals, we can conclude that they too had a similar auditory capacity to modern humans.

Unlike earlier anatomical developments relating to vocal communication, the enhanced motor control of the tongue and breathing, and the modernlike auditory capacity possessed by the Neanderthals, are most likely explained as having been specifically selected by evolution for vocal communication. The selective pressures for this are likely to have been numerous, including the need for enhanced communication with infants (Dissanayake 2000; and see below), hunting and gathering (Bickerton 2003), and the demands of living in larger and more complex social groups (Dunbar 1993).

Mithen (2005) followed Wray in proposing that protolanguage was holistic in nature and argued that early human vocalization was highly musical in character, making extensive use of variations in melody and rhythm to express emotion and induce emotional states in others. He suggested that early human vocal and gestural communication should be thought of as protomusic as much
as protolanguage. While “messages” would have consisted of arbitrary strings of syllables, early human utterances are also likely to have made extensive use of vocal imitation of natural sounds, especially animal calls, onomatopoeias, and to have had the phenomenon of sound synesthesia, as described by Berlin (2005). And as Donald (1991) argued, mimesis is likely to have played a key role in hominin communication.

However complex early human communication may have been, it seems unlikely to have possessed the key characteristics of modern day language, namely a set of grammatical rules that allow an infinite set of phrases to be generated from a finite set of words. Modern language generates creative thought, enables rapid cultural transmission, and cultural change through time (see below). These are precisely the features that appear absent from the archeological record of early humans. Their whole period of existence is characterized by immense cultural stability—a limited number of technological themes are simply used in different ways in different times and places. Other than a few highly ambiguous examples, there are no traces of symbolic behavior. Even with the big-brained Neanderthals, which have a fully evolved vocal tract, there are no traces of language-mediated thought and behavior (Mithen 2005). It seems most likely that their communication system was based on a suite of relatively fixed holistic phrases appropriate for frequently recurring situations (such as communal hunting, food sharing, toolmaking, social interactions) the emotional impact of which could have been changed by variations in pitch, rhythm, and melody (Mithen 2005).

### 22.4.2 Structure of mentality

When considering the cognition of early humans, paleoanthropologists have to engage with current (and long standing) debates as to the nature of the human mind/brain. Two of these are of primary relevance. First is the relative influence of genetic inheritance and cultural environment on the cognitive character of both species and the individuals of a species; second is whether mind/brains consist of general-purpose abilities, domain-specific modules, or some combination of both.

Evolutionary psychologists argue that whatever the process of development—whether by a neural Darwinism (Edelman 1987) or neural constructivism (Rosenzweig et al. 1999)—the modern human brain is constituted by a series of discrete modules rather than general-purpose learning abilities, explaining that this is the only structure of the brain that is compatible with evolution by natural selection (Cosmides and Tooby 1994; Pinker 1997). Modularity is, however,
conceptualized in various forms by different psychologists, some of whom do not align themselves with an evolutionary perspective. Fodor (1983), for instance, characterized the mind as constituted by modular input systems and a general-purpose central processor, while Gardner (1983) proposed that the mind/brain is formed by “multiple intelligences.” Karmiloff-Smith (1992) argued that modularity is a product of development, while Carey and Spelke (1994) argue the converse, suggesting that cross-domain mapping is the key character of cognitive development.

These debates about modularity were drawn upon by Mithen (1996) to suggest that early human mind/brains had a higher degree of modularity than that found in *H. sapiens*; indeed, he argued that the collapse of domain-specific thought, leading to what he termed “cognitive fluidity,” was the defining feature for the origin of modern humans. Mithen suggested that three principal cognitive domains constituted early humans minds. These he termed social, technological, and natural history intelligences. According to Mithen, each of these contained stores of knowledge and ways of thinking as complex as those found in modern humans and dedicated to that particular domain of activity. As a consequence, early humans were able to live in complex social groups with their social intelligence encompassing theory of mind abilities; similarly they were able to manufacture technically demanding stone artifacts such as Levallois points and handaxes by drawing on the understanding of fracture dynamics contained within their technological intelligence. Their survival in varying and often harsh landscapes of the Middle Pleistocene was dependent on their being expert folk botanists and zoologists, just as we find among modern hunter-gatherers, by virtue of having a “natural history intelligence.”

While the presence of these intelligences explains the cultural achievements of early humans, their isolation from each other explains the cultural stasis and absence of creative thought during the Middle and much of the later Pleistocene. The rather tedious nature of Middle Pleistocene culture—millennia after millennia of essentially the same types of artifacts and behavior—has often been remarked upon but been difficult to explain. Similarly, the absence of art, architecture, and specialized tools has also been difficult for paleoanthropologists to explain in light of the large brains possessed by early humans. Mithen (1996) proposed that such activities and cultural innovation in general depends on an integration of ways of thinking and stores of knowledge from more than one cognitive domain, and that this phenomenon was effectively absent in early human minds. Wynn and Coolidge (2004) suggest that the early human inability to achieve cognitive fluidity arose from constraints on their working memory, while Mithen (2005) follows Carruthers (2002) in attributing this constraint to the absence of compositional language, as further explained below.
A key to understanding the nature of the early human mind is the change in human life history that occurred during this period of human evolution (Bogin 2003). Modern humans contrast with apes by having infants that need far greater degrees of care, a distinct period of development that we term “childhood,” an adolescent growth spurt and menopausal life spans. Precisely when the shift from an apelike to a modern humanlike life history occurred remains unclear (Thompson et al. 2003). This was probably a gradual evolutionary process, intimately related to other evolutionary developments that included encephalization and bipedalism.

With the appearance of *H. ergaster*, male body size increased by up to 50% and female body size by up to 70% over their australopithecine forebears (Key and Aiello 1999). Brain size effectively doubled—that for australopithecines ranging between 400 and 500 cm³ and that for *H. ergaster* between 600 and 1,000 cm³. These increases in size would have put substantial energetic demands on pregnant and nursing females, while also inhibiting their ability to acquire food for themselves (Key and Aiello 1999). Such demands would have become particularly severe owing to the phenomenon of “secondary altriciality” that must have arisen at this time. This simply means giving birth to infants that are still undergoing fetal rates of development and consequently need constant care. This arose because bipedalism requires a relatively narrow pelvis and hence puts a severe constraint on the width of the birth canal. Human babies are literally forced along this and often have quite misshapen skulls for a period afterward. To be born at all through the narrow, bipedal pelvis, they had to be effectively born premature, leaving them almost entirely helpless for their first 18 months of life.

There are two consequences of secondary altriciality that require consideration. First is the impact on social relations. Owing to the increased energetic demands and constraints on movement of the pregnant and then nursing mothers, they would have required increased support for acquiring food and protection. Such support could have come either from other mature females within the group or from males who provisioned the mothers of their offspring.

The notion of mutually supportive female-kin networks has gained considerable popularity in recent paleoanthropological studies, partly through the studies of Aiello and Key, and partly due to the influence of the “grandmothering” hypothesis (Hawkes et al. 1997; O’Connell et al. 1999). This arose from fieldwork by the anthropologists James O’Connell, Kristen Hawkes, and Nicholas Blurton-Jones among the modern Hadza hunter-gatherers of East Africa. They found that although males spend a great deal of time hunting, they provide only limited amounts of food for the women and children. Of far
greater importance was the plant gathering by postmenopausal women, and the manner in which such women care for their grandchildren, enabling the physically stronger young mothers to forage for food. By providing such support, the grandmothers are improving their own inclusive fitness, although it should be noted that postmenopausal women in these communities also provide support to unrelated individuals. O’Connell et al. (1999) believe that this kind of female–female support is not only applicable to *H. ergaster* society but is directly related to the evolution of postmenopausal life spans within the *Homo* lineage—something quite absent in the African apes.

The grandmothering hypothesis suggests that the necessary support for pregnant and nursing *H. ergaster* females may have come from their female kin rather than their male mating partners. Would such female support have been sufficient? Key and Aiello (1999) attempted to estimate the increase in reproductive costs for females that arose from increased body and brain size during human evolution and assessed when it would have been to the advantage of males to begin provisioning for and protecting their mates. *H. ergaster* seems likely to have been just below a body/brain size threshold that marks a necessary transition from female–female to male–female cooperation. That most likely occurred at sometime between 500 and 100 ka when a further increase in brain size occurred (Ruff et al. 1997) and may reflect developments in hunting technology (Kohn and Mithen 1999).

Male provisioning of females and offspring was most likely a necessity among the Neanderthals owing to the high energetic demands on female reproduction arising from their large brain and body size. Demographics may also have played a role in changing the nature of male–female social relations. If Trinkaus (1995) is correct, and few Neanderthals survived beyond 35 years, there would have been a dearth of postmenopausal “grandmothers” to provide child support and provisioning to their daughters. Moreover, in light of the high mortality and injury rates among Neanderthals, it seems likely that there would have been a considerable number of orphans. It appears likely, therefore, that male Neanderthals would have engaged in provisioning females and offspring, which may have resulted in a form of serial pair-bonding.

A second consequence of relatively helpless infants arising from secondary altriciality is that there would have been enhanced selective pressures on mother–infant communications, which may have been important for the evolution of language. Dissanayake (2000) noted the highly musical nature of present day mother–infant interactions referred to as “motherese” or “infant directed speech” (IDS). She argued that such musicality arose because it provides substantial benefits to both parties by expressing and inducing emotional states that ultimately achieved a concordance of the emotions experienced by parent and
infant. Such concordance was essential, she argued, for developing their relationship and ultimately the enculturation of the infant: “by coevolution in infants and mothers of rhythmic, temporally patterned, jointly maintained communicative interactions that produced and sustained positive affect—psychobiological brain states of interest and joy—by displaying and imitating emotions of affiliation, and thereby sharing, communicating, and reinforcing them” (Dissanayake 2000 p 390).

Neanderthals may have been under particular selective pressures for the evolution of mother–infant vocal communication. Ramirez Rozzi and Bermudez de Castro (2004) compared dental enamel formation in the Neanderthals and modern humans and found that this was notably rapid in the former, indicating that they were the fastest developing of all species of Homo. Hence, the Neanderthals had a relatively shorter time to acquire their communication system, which for them meant learning a large number of holistic phrases, gestures, and body language and coming to understand the nuances of emotive expression.

Falk (2004) has also stressed the impact of secondary altriciality on the evolution of vocal communication but focused on the physical demands imposed on mothers of having to carry their infants around. Female monkeys and apes rarely put their babies down. Chimpanzee infants are unable to cling properly for the first 2 months of their life and are carried underneath their mothers’ bodies. When placed on the ground, infant chimpanzees are prone to emit a “hoo” sound as a means to reestablish physical contact with their mother. While such ape mother–infant turn-taking vocalizations are similar to those found in humans, they differ dramatically in their frequency and diversity. When compared with the constant sing-song chattering of a human mother, chimpanzees and bonobos are practically silent. When a couple of months old, chimpanzees and bonobos are sufficiently developed to climb onto their mothers’ backs and to cling onto their body hair for a free ride. They can do this for up to 4 years, after which the juveniles become entirely independent.

The period of physical helplessness of modern human infants lasts considerably longer; modern human infants also take longer than chimpanzees to gain control over their own posture and locomotion: approximately 3 months to lift their head alone and 9 months to sit unaided. A similar, although somewhat shorter, time span of physical incapacity is likely to have applied to the infants of our early human ancestors. As a consequence, the H. ergaster, H. erectus, H. heidelbergensis, and H. neanderthalensis mothers were required to support and carry their infants around for prolonged periods of time—unless they could find another strategy to care for them.

The vertical posture of bipedalism would have further increased the burden on mothers, as would have the large body size. We know that H. ergaster,
appearing at 1.8 Ma, had already reached modern human stature in some specimens while the Boxgrove leg bone and other skeletal remains indicate that by 0.5 Ma body size was often substantial—6-ft tall with plenty of muscle and fat. The infants would accordingly have been at least as large as those of modern humans, and hence carrying around a 1 year old would have been just as exhausting 1.8 Ma, as it was at Boxgrove, and as it is today.

It may have been even more so. It seems likely that hand in hand with the evolution of bipedalism came the loss of body hair, leaving just the few patches that we have today. Such loss would have been another physiological adaptation for keeping cool when foraging on the open savannah (Wheeler 1984, 1992). The early human infants are likely to have had a grasping reflex, but this would have become increasingly limited in value as their parents became less hairy. Although human infants are still born with a grasping reflex today, they never develop the ability to cling unaided to their mother's bodies.

How could mothers have both cared for their infants and foraged successfully? One possibility is by reliance on the “grandmothers” as discussed above for looking after the infants. This, however, leaves the problem of feeding. Unlike modern western mothers, hominins most likely provided on-demand feeding as is pervasive in traditional societies. A second possibility is that the babies were “parked”—left in a safe place such as the crown of a tree—while the mother undertook her foraging, toolmaking, courtship, and other activities at some distance from the parking spot, and perhaps for lengthy periods of time. This is a tactic used by many small primates such as langurs and lemurs. But it is very rarely seen in the higher primates. This is not surprising as it risks having one's infant injure itself, killed by predators, or even killed by another member of one’s own group (such as a male who is not the father). Also, because infants require frequent feeding, lactating mothers would have been constrained as to the distance they could travel from the “parking spot.” In general, the hunter-gatherer-scavenger lifestyle on the Africa savannah, with intense social competition and prowling carnivores, is the exact converse of that amenable to “baby parking.”

A third alternative to the big, helpless, demanding baby problem is the frequent “putting down” of the infant for short periods within sight and sound of the mother—unlike “parking” in which the possibility for such contact is lost. This would allow the mother to use both hands to pick fruit, butcher a carcass, drink from a river or knap stone; without the baby she could reach, bend, run, and stretch, and simply gain some relief from carrying its weight around. When she was ready to move on, perhaps to a new patch of berries or to use a newly chipped stone flake, she could simply scoop baby up from the ground, as she could whenever the child became distressed.
Falk (2004) suspects that such “putting down” had occurred and was essential to the development of “prelinguistic communication.” For once the baby is put down, the mother would have eye-contact, gestures, expressions, and utterances to reassure the infant, these being a substitute for the physical contact that the infant desires. The emotionally manipulative prosodic utterances that constitute IDS would, Falk suggests, have been a “disembodied extension of the mother’s cradling arms.” In this regard, a *H. ergaster* or *H. heidelbergensis* mother would have been little different from a *H. sapiens* mother today.

In summary, Falk argues that those mothers who had a biologically based propensity to attend to their infants when “put down,” by making use of vocalizations, expressions, and gestures, were strongly selected—their genes, and hence such behaviors, would have spread in the population. She argues that the IDS that evolved was initially prelinguistic and constituted by the types of melodic and rhythmic utterances that lack symbolic meaning but which have the emotional. But over time, according to Falk, “words would have emerged in hominids from the prelinguistic melodies and become conventionalised.”

### 22.4.4 Summary: problematic societies of early humans

The hominins considered in this section were large brained, fully bipedal, and lived in diverse environments throughout much of the Old World. Generalizations are risky because there was considerable variation in body and brain size throughout this period; indeed, it encompasses the critical time between 600 and 100 ka when brain size shows a quite astonishing period of enlargement.

One of the key problems when interpreting the fossil and archeological records of these hominins is the absence of any analogous species in the modern world. When studying the australopithecines and earliest *Homo*, one can feel confident on drawing on comparative socioecology of living primates—the patterns we find today between body and brain size, between environment, group size, and diet among apes and monkeys—to inform our interpretations of the past. Similarly, when considering the earliest *Homo sapiens*, as I will do so in my next section, we can cautiously draw on studies of modern hunter-gatherers to facilitate our studies. But species such as *H. ergaster*, *H. erectus*, and *H. heidelbergensis* lack any suitable modern-day analogy.

These species appear in some ways to have been very similar to modern day humans while in others to have been radically different. This is especially the case of the Neanderthals who had equivalently large brains were highly skilled tool-makers and big game hunters and yet who had a relatively static culture across thousands of years and dramatic episode of environmental change. It seems most
likely that they had a sophisticated form of spoken communication and yet this appears to have fallen short of language (Mithen 2005). What is clear is that when we try to understand their societies and that of their immediate ancestors, we have to explore the network of links between evolving body and brain size, life history, communication, and the structure of mentality.

22.5 Origins of modern humans and modern minds

*H. sapiens* evolved in Africa, most likely around 200 ka, a date on which both the fossil and genetic evidence increasingly converge (Ingman et al. 2000; McBrearty and Brooks 2000; White et al. 2003; Jobling et al. 2004; McDougall et al. 2005). It is not until 70 ka, however, that we find unambiguous traces of symbolic activity (McBrearty and Brooks 2000; Henshilwood et al. 2002, 2004). This lapse in time might be a consequence of our limited knowledge of the African archeological record in the critical intervening period. Alternatively, 70 ka may mark a date at which human populations passed a demographic threshold that enabled enhanced cultural activity (Shennan 2000) or when a further genetic mutation happened that changed the nature of human mentality by enabling cognitive fluidity. In the context of this essay, there are two key issues to address regarding the network of brain, body, language, and culture in modern humans: the evolution of language and the “extension of mind.”

22.5.1 Segmentation of holistic protolanguage and cognitive fluidity

In this section, I will assume that the protolanguage of early humans, including the immediate ancestor of *H. sapiens* in Africa, was holistic in nature. Wray (1998, 2000) uses the term “segmentation” to describe the process whereby hominins began to break up holistic phrases into separate units, each of which had its own referential meaning and could then be recombined with units from other utterances to create any infinite array of new utterances. This is the emergence of compositionality, the feature that makes language so much more powerful than any other communication system.

Wray suggests that segmentation may have arisen from the recognition of chance associations between the phonetic segments of the holistic utterance and objects or events to which they related. Once recognized, these associations might then have been used in a referential fashion to create the new, compositional phrases. Bickerton (2003) and Tallerman (in press) question the feasibility of
“segmentation” as the process by which words were invented/discovered. They have several pointed criticisms: the likelihood of such chance associations arising are remote; if holistic utterances had existed as Wray describes, then they are unlikely to have been long enough to contain multiple phonetic segments that could have been partitioned in this fashion; Wray’s theory of segmentation has to assume the prior existence of discrete segments which then invalidates her conception of holistic utterances.

Such criticisms do not take into account the likely presence of onomatopoeia, vocal imitation, and sound synesthesia in early human holistic utterances. These would have created nonarbitrary associations between phonetic segments of holistic utterances and certain entities in the world, notably species of animals with distinctive calls, environmental features with distinctive sounds, and bodily responses (Mithen 2005). These nonarbitrary associations would have significantly increased the likelihood that the particular phonetic segments would eventually come to refer to the relevant entities and hence exist as words. Once some words had emerged others would have followed with greater ease by the segmentation process that Wray describes.

Further confidence in the process of segmentation comes from the use of computer models to simulate the evolution of language. Kirby (2000, 2002) is one of several linguists who have begun to explore the evolution of language by the use of computer simulation models (Nowak and Komarova 2001; Batali 2002; Komarova and Nowak 2003). Kirby creates populations of agents—simulated people—which communicate with each other by using strings of symbols for language. Each simulated population is multigenerational, with each new “learning-agent” acquiring its language from the already existing “speaking-agents.” In general, Kirby’s simulations indicate that grammatical structure can be an emergent phenomenon of the cultural transmission of language between generations.

While the arguments of Wray and Kirby help us to understand how compositional language evolved from holistic phrases, we must ask why this appears to have only happened in Africa after 200 ka. There are two possibilities, one relating to social life and one to human biology. With regard to the first, we should initially note that Kirby found holistic languages remain stable in those simulations in which learning-agents hear so much of the speaking-agents’ utterances that they learn every single association between symbol-string and meaning. In other words, there is no learning bottleneck for language to pass through and hence no need for generalization. It may be the case, therefore, that the social arrangements of Homo populations other than H. sapiens were such that infants had intense and continuous exposure to a limited number of holistic speakers, resulting in the acquisition of a whole suite of utterances intact with no need for generalization.
This would indeed have been quite likely for early humans as they most likely lived in socially intimate groups with limited, if any, need for the type novel utterances that could only be produced by compositional language. Moreover, there would have been little need and few opportunities to communicate with people from outside one’s own group—although some contacts involving the movement of people would have been essential to maintain demographic and genetic viability. But little need have been said on such occasions.

It may only have been within the earliest *H. sapiens* communities in Africa that people began to adopt specialized economic roles and social positions that trade and exchange with other communities began and that “talking with strangers” became an important and pervasive aspect of social life. Such developments would have created pressures to exchange far greater amounts of information than was previously necessary in the socially intimate, undifferentiated groups of early humans. Only then would there have been the need for generalization, in the manner that Kirby describes within his simulation, and to continually generate novel utterances at a rate and of a type beyond the capability of holistic protolanguage.

The archeological record of Africa relating to the appearance of *H. sapiens* certainly suggests that such social developments occurred (McBrearty and Brooks 2000). The dilemma, of course, is whether we are dealing with cause or effect: one might argue that the development of economic specialization and exchange relations between groups were a consequence of compositional language that enabled the necessary communication to be undertaken. My guess is that we are dealing with strong feedback between the two—they “boot-strapped” each other to create rapid changes in both society and communication.

The kick-start to such developments may have been a chance genetic mutation. This may have provided the ability to identify phonetic segments in a holistic utterance, an ability that had previously been absent. Some aspects of language are dependent on the possession of the specific gene FOXP2, the modern human version of which seems to have appeared in Africa at soon after 200 ka (Enard et al. 2002). Perhaps the process of segmentation was dependent on this gene in some manner that has yet to be discovered. Indeed, it may be significant that those members of the KE family who were afflicted by a faulty version of the FOXP2 gene not only had difficulties with grammar but also with understanding complex sentences and judging whether a sequence such as “blontesterpating” is a real word (Bishop 2002). These difficulties seem to reflect a problem with the segmentation of what would have sounded to them as holistic utterances. So perhaps it was only with the chance mutation of the FOXP2 gene to create the modern human version that segmentation became possible. Alternatively, there may have been other genetic mutations at a similar date that enabled
the transition from holistic phrases to compositional language, perhaps by the appearance of a general-purpose statistical learning ability.

Once the process of segmentation had begun, we should expect a rapid evolution of grammatical rules. Such rules would have evolved via the process of cultural transmission in the manner that Kirby describes, and perhaps through natural selection leading to the appearance of genetically based neural networks that enable more complex grammatical constructions. While the evolution of compositional language in this manner would have enhanced communication, it may have also led to a more fundamental change in the nature of human thought—the transition from a domain—specific to a cognitively fluid mentality.

Mithen (1996) argued that cognitive fluidity was a consequence of language: spoken and imaginary utterances acted as conduits for ideas and information to flow from one intelligence to another, a view supported by Carruthers (2002). He drew on latest research in neuroscience and psychology to argue that the “imagined sentences” we create in our minds allow the outputs from one intelligence/module to be combined with those from one or more others, and thereby create new types of conscious thoughts. Carruthers placed considerable emphasis on syntax—an essential part of compositional language. Syntax allows for the multiple embedding of adjectives and phrases, the phenomenon of recursiveness. According to Carruthers, syntax allows one imaginary sentence generated by one type of cognitive module/intelligence to be embedded into that of another imaginary sentence coming from a different module/intelligence. By so doing, a single imaginary sentence will be created which generates an “intermodular” or cognitively fluid thought, one that could not have existed without compositional language.

Any cognitive impact that language has on an individual is magnified a multitude of times by the manner in which language connects people’s minds into a single network, allowing ideas to “migrate” between minds. Indeed, language creates a communal mind/brain. The philosopher Andy Clark (1996 p 206) explains this remarkable power of language:

▶ Migrations [of ideas by language] may allow the communal construction of extremely delicate and difficult intellectual trajectories and progressions. An idea that only Joe’s prior experience could make available, but that can flourish only in the intellectual niche currently provided by the brain of Mary, can now realise its full potential by journeying between Joe and Mary as and when required. The path to a good idea can now criss-cross individual learning histories so that one agent’s local minimum becomes another’s potent building block . . . culturally scaffolded reason is able to incrementally explore spaces which path dependent reason could never hope to penetrate.
22.5.2 Material culture and the extension of mind

The evolution of compositional language provided *H. sapiens* with enhanced communicative abilities and changed the capacity for thought by enabling cognitive fluidity. One consequence of this was a change in the relationship with material culture, one that had a profound influence in the development of human history. The use of material culture had, of course, been a fundamental aspect of hominin lifestyles for millions of years, probably long before the first stone artifacts were manufactured in light of the use of tools by African apes. But until 70 ka—on the basis of current evidence—we have limited if any evidence that material culture was used in a manner other than to extend the physical capabilities of the human body. There are no unambiguous traces of artifacts that were used for social communication, that may have carried symbolic meanings or which enhanced the computational powers of the human mind. Today, of course, such artifacts are pervasive in all of our activities ranging from the “post-its” that extend our memories, calculators that extend our computational powers, and symbolic images that extend our abilities for metaphorical thought. Our reliance on technology is now so great that Clark (2003) recently described *H. sapiens* as “natural born cyborgs.” The development of this new type of network between brain and material culture began with the origin of modern humans.

The emergence of cognitive fluidity lies at its root as this both created the need for material culture to play this role and enabled the necessary technological creativity for it to do so. The development of religious thought, involving a belief in supernatural beings, provides a particularly good illustration.

Ideas about supernatural beings are a “natural” consequence of the cognitive fluidity that language delivered to the human mind (Mithen 1996). By combining, say, what one knows about people (from social intelligence) with what one knows about animals (from natural history intelligence), one can imagine an entity that is part human and part animal such as the lion/man figurine from Hohlenstein-Stadel in Germany dating to 33 ka. Similarly, by combining what one knows about people with what one knows about physical objects, one can create ideas of beings that can live forever (like rocks), walk on water (like a floating branch), or are invisible (like the air we breath).

Ideas about supernatural beings are the essence of religious thought (Boyer 2001), but they are unnatural in the sense that they conflict with our deeply evolved domain specific understanding of the world. As a consequence, they are difficult to hold within our minds and transmit to others—try, for instance, explaining to someone the concept of the “Holy Trinity” or the “Aboriginal Dreamtime” or try understanding these ideas as someone explains them to you. As Day (2004 p 116) has recently written “one of the bedevilling problems about
dealing with gods is that ... they are never really there” and hence we have difficulty in knowing not only how to communicate with them but also how to think about them.

Modern humans compensate for this by the use of material symbols that provide “cognitive anchors” for such difficult ideas, whether about supernatural beings or scientific theories (Mithen 1998a). Whether supernatural beings are made tangible in a representational manner, as we suppose the lion/man from Hohlenstein-Stadel is doing, or in abstract form as in the Christian Cross, such material symbols function to help conceptualize and share the religious entities and ideas that one believes. In this regard, such objects constitute an extension of the human mind. Music can also act as a cognitive anchor and extend the capacities of the mind. Just as a community can share a nondiscursive religious idea by all wearing the same material symbol, they can do so by all making the same music. In this regard, music plays an active role in maintaining and manipulating ideas about supernatural beings in the human mind; without such support from music and material symbols, such ideas would simply be too difficult to think about and share other than for a fleeting instance. As Day (2004 p 116) explains:

- the broad spectrum of rituals, music, relics, scriptures, statues and buildings typically associated with religious traditions are no longer seen as mere ethno- graphic icing on the cognitive cake. Rather than thin cultural “wrap arounds” that dress-up the real cognitive processes going on underneath, they begin to look like central components of the relevant machinery of religious thought.

The use of material culture and music to extend the capabilities of the human mind by providing anchors for ideas about supernatural beings that would be otherwise difficult to maintain, comprehend, and transmit is just one example of a much more general phenomenon. We can, for instance, consider how the use of shell beads, such as those dating to 70 ka at Blombos Cave (Henshilwood et al. 2004), and other forms of body ornamentation, function to mediate social relationships by indicating social identity and/or status. Rock art enables information and ideas to be recalled and communicated across time and space far in excess of what a human body and brain are able to do alone. Indeed, paintings from more than 30 ka are still communicating ideas to us today. The invention of pictograms and writing provide a means by which narratives that have a level of detail beyond that which can be depicted in art alone can be recorded, communicated, and analyzed in a manner that is not possible with oral recitation alone. The invention and development of mathematics and science are entirely dependent on off-loading information and ideas into material culture; these can then be examined further and developed by either the same mind or a different mind at a different time and place.
In this regard, the quotation from Clark given above concerning language can also apply to material culture; in fact, it is even more appropriate for material culture (Mithen 1998b). While language is effective at allowing ideas to migrate between minds, it has a major constraint—utterances are not durable. If there is no one around to hear what one says, the idea within the utterance is lost or at least remains within the mind of the speaker. But once ideas are encoded into durable media, they become part of the world and can be communicated across vast time spans. As such, material culture is the prime means by which minds are extended beyond the body and connected to other individuals who may never meet each other or even know of each other’s existence.

22.5.3 The dispersal and cultural evolution of modern humans

The cognitively fluid and extended mind of *H. sapiens* led to radically different types of behavior from all hominin predecessors. This becomes manifest with the dispersal into Europe, Southeast Asia, and Australia soon after 50 ka, the appearance of representational art in the northern and southern hemispheres, and a whole wealth of technological innovations that are collectively referred to in Europe as the transition to the Upper Paleolithic. All other species of hominins became extinct, notably the Neanderthals in Europe, although *H. floresiensis* appears to have survived on an isolated island in Indoensia until 13,000 BC.

The most striking difference between *H. sapiens* and their ancestors and relatives only became apparent with the start of the last interglacial at around 10,000 BC. Whereas during all previous interglacials hominins had simply responded by extending or retracting their foraging range and adjusting their hunting and gathering strategies, *H. sapiens* engaged in radically new behavior: farming. This was not only a new economic strategy but also involved sedentism which provided the possibility for a massive investment in new forms of material culture ranging from monumental architecture to pottery. And such material culture provided a host of new ways to extend cognitive abilities, culminating in the invention of writing at around 3,000 BC.

22.6 Finis: the accumulation of networks of brain, body, language, and culture

During the course of this chapter, I have stressed the significance of different networks between brain, body, language, and culture at different time periods of
human evolution. For the australopithecines and earliest Homo, I focused on that between environment, diet, bipedalism, and sociality; for early humans the emphasis was on language, mentality, and life history, while for H. sapiens I have stressed the role that material culture plays in extending both physical and mental abilities. This is, of course, a simplification, as throughout human evolution all of these networks have been active at all times and perhaps the only key trend is that they have become increasingly complex and intertwined. Hence, although I did not discuss diet with regard to H. sapiens, we know that the nature of diet plays a critical role in cognitive and physical development; likewise, material culture in the form of Oldowan tools was an essential element of the network that related foraging, bipedalism, and brain size. It is the paleoanthropologist’s task to untangle these networks and understand the coevolution of body, brain, language, and culture.

References


Boesch C (1991) Teaching among wild chimpanzees. Anim Behav 41: 530–532


Lordkipanidze D, Bar-Yosef O, Otte M (2000) Early humans at the gates of Europe. ERAUL 92, Liège
Tallerman M (in press) Analysing the analytic: Problems with holistic theories of protolanguage
Abstract

This chapter examines large-scale patterns of behavioral change that are often viewed as indicators for the advent of cultural modernity and developed symbolic communication. Using examples from Africa and Eurasia, the chapter reviews patterns of lithic and organic technology, subsistence, and settlement as potential indicators of modern behavior. These areas of research produce a mosaic picture of advanced technology and behavioral patterns that come and go during the late Middle and Late Pleistocene. Based on these data the emergence of modern behavior, as seen in the archeologically visible material record, appears to be gradual and heterogeneous in space and time. During the early part of the Late Pleistocene, personal ornaments in the form of sea shells are documented in southwestern Asia and southern Africa. By about 40 thousand years ago (ka), a diverse array of personal ornaments is documented across the Old World in association with Neanderthals and anatomically modern humans in Europe. These include both modified natural objects and fully formed ornaments. The timing and distribution of the appearance of figurative art and other classes of artifacts including musical instruments point to a more punctuated development of fully modern behavior during the middle of the Late Pleistocene, and certainly no later than 40 ka. Due perhaps in part to the long and intense history of research much, but by no means all, of the relevant data come from Europe. Early figurative art from the Aurignacian of southwestern Germany, northern Italy, Austria, and southern France provides undisputed evidence for fully developed symbolic communication and behavioral modernity. This chapter also discusses some of the hypotheses for the development and spread of cultural modernity and rejects a strict monogenetic model in favor of a pattern of highly variable, historically contingent, polygenetic development. This chapter highlights the need for new refutable, regional and superregional hypotheses for the advent and spread of behavioral modernity.
23.1 Introduction

The question of when in the course of human evolution hominids became like ourselves has been at the center of several decades of productive debate in paleoanthropology. Reduced to the most fundamental level, the appearance of anatomical and behavioral modernity is a question of at what time in the course of evolution hominid anatomy and behavior fall within the variability documented in recent societies. The key component of fully modern cultural behavior is communication within a symbolically organized world and the ability to manipulate symbols in diverse social contexts.

This chapter will not address the development of modern human anatomy; here, I consider some of the key evidences for the evolution of complex behavioral systems. While there is no consensus about when modern behavior can first be identified in the archeological record, by no later than about 40 ka diverse finds documenting the production of ornaments, musical instruments, and figurative art provides undisputed evidence for cultural modernity. These and other archeologically visible indicators of cultural modernity point to a patchy development of complex cultural behavior and symbolic communication across the Old World. While some regional patterning is becoming visible (Delporte 1998; McBrearty and Brooks 2000; Bon 2002; Conard and Bolus 2003), the current data on this topic are generally a hodgepodge of evidence that has been put through a selective taphonomic filter and reflects diverse regional histories of research. These biases hinder the location of convincing centers of origin and dispersal for many of the key features considered here.

At present, we see diverse points of view regarding the origins of behavioral modernity, and current interpretations include but not limited to the following models are: (1) gradual African origin (McBrearty and Brooks 2000), (2) coastal origin in connection with new dietary patterns during the early Late Pleistocene (Parkington 2001), (3) punctuated late African origin (Klein 1999; Klein and Edgar 2002), (4) gradual origins across multiple human taxa and multiple continents (d’Errico et al. 2003), and (5) relatively late origins among multiple human taxa, including “Neanderthals’ own Upper Palaeolithic revolution” (Zilhão 2001 p 54). Here, I argue for gradual polygenetic origins of behavioral modernity within a dynamic equilibrium between anatomically archaic and modern humans. The evolution toward behavioral modernity accelerated in the middle of the Late Pleistocene, and culturally modern behavior with diverse regional signals and local innovations, can be seen in many parts of Africa, Europe, Asia, and Australia between 30 and 40 ka. While archaic and modern humans must have interacted in many regions in the context of diverse social and ecological
settings, ultimately modern humans were at a demographic advantage in all regions and replaced archaic humans with relatively little interbreeding.

This chapter reviews some of the evidence for advanced cultural behavior and argues for a highly variable pattern of development depending on specific historical and evolutionary contingencies. The development of modern behavior does not in my view represent a one-time-only quantum leap, but a complex pattern of innovation, spread, and local extinction of new traits through cultural selection and social reproduction. Social, technological, and linguistic reproduction through learning are fostered by the biological success of the members of societies, but are not only driven by demographic growth. Demographic trends and complex patterns of intra- and intersocietal contacts led to mosaic patterns of cultural development that result from specific historical and ecological occurrences during the Pleistocene. The current archeological record provides glimpses of these evolutionary processes, but it would be naive to think that our current data on the fleeting material remains of the development and spread of behavioral modernity provide a one-to-one indication of where and when advanced technology, highly developed patterns of settlement and subsistence, ornaments, music, abstract and figurative representation evolved. The question of why fully modern cultural behavior evolved is still more difficult to answer, but recent years have begun to see attempts to address the thorny questions of causality (Klein 1999; Parkington 2001; Shennan 2001; Lewis-Williams 2002; Conard and Bolus 2003). Much more work is needed that addresses the potential causes of cultural evolution and develops testable hypotheses. In this context, monogenetic and polygenetic models need to be formulated and tested explicitly.

Turning to the more mundane aspects of archeology, it is necessary to stress the ambiguities and problems with dating sites in excess of 30,000 years. Radiocarbon dating, the strongest tool for dating LSA and Upper Paleolithic assemblages, begins to reach its limits in the period before 30 ka. Here, several factors come into play. In this period, in excess of five radiocarbon half-lives, contamination becomes a serious problem. The isolation of preserved collagen in bones and similar problems related to sample preparation become more problematic than in younger periods. Also, the physics of the AMS and beta counting become more challenging as minimal contamination begins to affect the results more strongly and the uncertainties related to the chemistry and instrument background become significant. Equally important is the wealth of evidence that there are major fluctuations in radiocarbon levels, probably in connection with variations in production due to magnetic excursions (Voelker et al. 2000; Beck et al. 2001; Conard and Bolus 2003; Hughen et al. 2004). These factors tend to make
radiocarbon measurements underestimate the calendar age of archeological materials in excess of 30,000 years.

Other methods, including luminescence dating, have great potential for sorting out the chronology of the emergence of modern human anatomy and behavior, but have yet to find broad application in addressing these issues. Fortunately, this situation is improving rapidly, and the prospects for gaining improved chronological control for the later stages of human evolution are excellent (Richter et al. 2000; Jacobs et al. 2003a, b).

This presentation will of necessity be brief and in no way attempts to be encyclopedic. Instead, I consider examples to illustrate the overall pattern of behavioral evolution. These examples are often drawn from regions where I have worked and know the data best. The subject matter is divided into two main sections. The first deals with the nuts and bolts of Paleolithic archeology and focuses on lithic and organic artifacts and patterns of subsistence and settlement. The second section deals more with data that provide more direct access to Paleolithic world of symbols, beliefs, and communication, and reviews evidence for burials, ornament, figurative, and non-figurative representation and music as means of defining modern cultural patterns. In general, the results from a review of the latter kinds of evidence give a better indication of the origins of behavioral modernity. My concern here is not in developing trait lists or single signatures for modernity, but rather to look at the evolutionary contexts of diverse classes of data that may help us to identify patterns of behavioral evolution. Other similar reviews of this evidence at different geographic scales can be found in a number of recent publications and should be consulted along with the primary references for further details (Deacon and Deacon 1999; Klein 1999; McBrearty and Brooks 2000; d’Errico 2003; d’Errico et al. 2003; Conard 2004a). Finally, many of the chapters in this volume present up-to-date information that is of central importance for defining the evolution of modern behavior.

23.2 Technology, settlement, and subsistence as measures of modernity

23.2.1 Lithic technology

Stone artifacts are a physically robust class of artifacts and often survive the numerous potential forms of taphonomic destruction. In this regard, they are a major source of data on early human behavior. In many Paleolithic settings, stone artifacts are the most abundant class of anthropogenically altered material. These attributes of lithic artifacts make them the most important means of defining
Paleolithic cultural groups. Thus, if specific lithic artifacts provided an indication of modern cultural behavior, scholars could use such finds as indicators of modernity.

Despite attempts to define linear or even cladistic systems for the evolution of stone tools (Foley 1987; Foley and Lahr 2003), lithic technology is based on learned behavior and is not directly transmitted biologically. Thus, it comes as little surprise that new forms of lithic technology come and go over the more than 2-million year old Paleolithic record. Oldowan technology is the most common form of flint knapping at the pyramids of Giza (Conard 2000), and this simplest of knapping approaches comes and goes throughout the Stone Age. Many other knapping technologies also come and go over the last several hundred thousands of years that are the backdrop for the development of anatomical and cultural modernity. Handaxes, Levallois technology, blade technology, and other elements of stone knapping come and go and do not provide certain indicators of modernity.

Additionally, the ethnographic record points to the problems associated with viewing lithic technology as a clear indicator of levels of cultural evolution. Subrecent ethnographic sources document cases of hunter-gatherers in regions including, for example, parts of Australia and Tierra del Fuego, who used Stone Age technologies that would leave no traces of behavioral modernity. These groups were undeniably modern humans and highly developed in respect to their linguistic skills and their ability to manipulate symbols, yet the lithic technology and the archeologically visible material culture would leave no traces of this modernity.

Lithic technology provides no simple solution to the problems related to identifying modernity. Even blades, which were once seen as clear indicators for behaviorally modern, Upper Paleolithic and Later Stone Age cultures have been demonstrated in diverse contexts in Africa, the Near East, and Europe (Rust 1950; McBurney 1967; Besançon et al. 1981; Singer and Wymer 1982; Conard 1990; Révillion 1994; McBrearty and Brooks 2000) (Figure 23.1). These blade-based assemblages date to the second half of the Middle Pleistocene and the Late Pleistocene, and include technologies based on Upper Paleolithic platform cores, non-Levallois and Levallois blade production.

Lithic assemblages document a heterogeneous pattern of development with forms coming and going across the Old World. While in Europe there is no doubt a difference between Middle and Upper Paleolithic assemblages, many forms typically associated with the Upper Paleolithic appear in earlier periods, and it is becoming increasingly clear that the variability documented by Bordes (1961) in the Middle Paleolithic of southwestern France reflects only a small portion of the overall lithic variability. Many regions of Europe (Bosinski 1967, 1982; Conard and Fischer 2000) show a diverse pattern of cultural development that
is analogous to that documented in Africa (Clark 1982, 1988; McBrearty and Brooks 2000). Also in the Near East, the early Middle Paleolithic includes lithic assemblages such as Yabrudian and Humallian, and the later Middle Paleolithic is characterized by Levalloisian assemblages that were made by both Neanderthals and anatomically modern humans (Shea 2003). The latter observation demonstrates how tenuous the link is between anatomical and cultural evolution.

As Bosinski (1982), Clark (1982, 1988), and others have long pointed out, the MSA and Middle Paleolithic are marked by the growth and increased visibility of local traditions. The frequently made suggestion that lithic technology from these periods is static or even boring, strike me as incorrect. In many areas where high quality data are available, MSA and Middle Paleolithic assemblages show considerable diversity. The development of local traditions appears to increase with time in some areas of Africa and Eurasia (Bosinski 1967; Conard and Fischer 2000; Wadley 2001; Jöris 2002), but these trends are, to a certain extent, a reflection of the improved quality of data that results from both better chronological control and more numerous assemblages per unit time. Researchers who try to define variability must consider the quality and density of the available data. In general, early periods of the MSA and Middle Paleolithic have provided less data to address these questions than the later phases of these periods or the LSA or Upper Paleolithic. Thus it is not surprising that, in general, assemblages from more recent periods document more technological and typological variation than samples from earlier periods.

The complexity of Middle Paleolithic and MSA lithic technology is highly variable, but at times advanced. Hafting and composite tools have been documented...
directly and indirectly in many regions. In Africa, we can consider the standardized-backed forms from Howiesons Poort assemblages to be strong candidates for hafting, as well as numerous point assemblages of the Upper Pleistocene and perhaps the Middle Pleistocene (Singer and Wymer 1982; McBrearty and Brooks 2000) (Figure 23.2). In southwestern Asia, Shea (1988, 1993, 1998) has long

Figure 23.2
Klasies River Mouth, South Africa. Highly standardized lithic artifacts from the Howiesons Poort assemblage ca. 75,000 years old (after Singer and Wymer 1982)
argued for hafting based on patterns of damage to artifacts and use wear. Mastic attached to Middle Paleolithic artifacts at Umm-el-Tlel in central Syria also demonstrates the use of hafting and provides evidence for composite tools (Boëda et al. 1998). In Europe, a similar pattern is present with small-backed artifacts that almost certainly required hafting being recovered at Tönchesberg (Conard 1992). European chipped stone points would have required hafting as on other continents, and mastic has been recovered, for example, at Middle Paleolithic sites of Königsau (Mania and Toepfer 1973) Neumark-Nord (Mania et al. 1990; Meller 2003) and Bocksteinschmiede (Wetzel and Bosinski 1969) in Germany.

Lithic assemblages of the MSA and Middle Paleolithic do not provide the evidence needed to define precisely when modern patterns of human behavior developed. They do, however, clearly show a heterogeneous pattern of technological development and transmission that does not indicate that the beginnings of the LSA and Upper Paleolithic saw fundamental revolutionary changes in technology across the Old World. This transition saw change and the further development of new technologies, but while more advanced forms of lithic technology came into broader use in the LSA and Upper Paleolithic, most of these technologies have well-documented precursors in earlier periods.

23.2.2 Organic technology

The development of organic technology shows a pattern analogous to that of lithic technology. While the LSA and Upper Paleolithic are defined on the basis of new artifact forms that occur in easily detectable numbers, organic artifacts have antecedents extending into the ESA and Lower Paleolithic. Thus the beginnings of the LSA and Upper Paleolithic reflect legitimate archeological divisions, but the changes represent a further elaboration and intensification of technologies that in some cases existed earlier.

In regard to this question, the most important discoveries of the last decade are the finds from Schöningen in northern Germany, where Thieme’s excavations have yielded eight wooden spears and numerous other wooden tools (Thieme 1997, 1999) (Figure 23.3). These tools are of the highest workmanship and lend support to the importance of wooden tools from Clacton-on-Sea (Oakley et al. 1977) and Lehringen (Thieme and Veil 1985). Unless we postulate that this part of eastern Niedersachsen enjoyed a privileged position in human cultural evolution, we must conclude that organic technology and diverse well-made wooden tool assemblages were a part of daily life of the Lower and presumably Middle Paleolithic inhabitants of Europe. These sites provide a highly favorable setting for preservation that cannot be matched in other sedimentary settings, but
Figure 23.3
Tönchesberg 2B, Germany. Middle Paleolithic assemblage with blades, bladelets, backed points and backed bladelets and imported lithic materials ca. 100,000 years old (after Conard 1992)
occasional finds of preserved wood in Africa and the Near East leave room for optimism that future work may uncover comparable wooden artifacts.

Much has been made of the development and elaboration of bone, ivory and antler tools in recent years (Gaudzinski 1999; d’Errico 2003; d’Errico et al. 2003). MSA assemblages from sites including Apollo 11 (Vogelsang 1998), Klasies River (Singer and Wymer 1982), and Blombos (Henshilwood et al. 2001) have pro-

![Figure 23.4](image)

Schöningen, Germany. Lower Paleolithic wooden spear and horse bones ca. 350,000 years old (photo N.J. Conard)

duced a wealth of bone artifacts (**Figure 23.4**). Many examples are sharpened bones and bone splinters. Other bone tools show series of notches or more enigmatic forms. An exceptional case is the elaborately made harpoons from Katanda in D. R. Congo, former Zaire (Brooks et al. 1995); these finds would be remarkable if they were indeed of early Late Pleistocene age. Certainly, by the middle of the Late Pleistocene simple bone tools were widespread in the MSA.
The European Lower Paleolithic also documents early examples of bone tools including carefully manufactured handaxes (Segre and Ascenzi 1984; Gatti 1993). Similarly, bone tools are well documented at Middle Paleolithic sites, including Salzgitter-Lebenstedt (Gaudzinski 1999), Große Grotte (Wagner 1983), and Vogelherd (Riek 1934). Bone tools are by no means as common or complex as those of the Upper Paleolithic, but they no doubt existed in Middle Paleolithic assemblages. Bone tools were clearly used by late Neanderthals in many settings, and they have occasionally been documented in large numbers (d’Errico et al. 2003). These tools tend to be less standardized and less elaborate than the organic tools of the Aurignacian. Here, split base points, for example, are marker artifacts for the early Aurignacian over much of Europe (Albrecht et al. 1972; Hahn 1977). These standardized forms occur in Europe in significant numbers starting around 40 ka.

Finally, the Late Pleistocene sees further evidence for cultural innovations that should be mentioned here. These innovations include the widespread use of grinding technology during the MSA and Middle Paleolithic of northern Africa (Wendorf et al. 1993; Van Peer et al. 2004), evidence for fire-making technology in the Swabian Aurignacian (Riek 1934: 161; Weiner and Floss 2004), and water transport technology in the form of perforated ostrich eggshells (Vogelsang 1998; Parkington et al. 2005; but see Kandel 2004). As these and other less well-studied categories of finds and behavioral innovations become topics of more systematic research, they will play a more prominent roll in the discussions about the evolution of cultural modernity.

### 23.3 Subsistence

Patterns of subsistence vary in time and space due to changing environmental conditions and changes in technology combined with changing social and settlement strategies. Although most sites do not contain preserved botanical remains, there is every reason to assume that plants played an important part in the diet of all hominids, just as they do for all ethnographically documented societies (Owen 2005). The diet of Neanderthals as reflected in stable isotope data indicates a relatively high component of animal resources (Bocherens et al. 1999, 2001), but these results do not preclude the use of plants in the diet and even in the harshest arctic and desert environments, plants are seasonally available and nutritionally important.

This is not the appropriate place to summarize the history of research on this question, but recent decades have seen a shift from assuming that archaic and early modern humans practiced fully developed systems of hunting and food sharing to a critical assessment and rejection of the earlier interpretations by many Anglophone colleagues. More recently, many case studies have provided
convincing evidence that both later archaic and anatomically modern humans practiced systematic hunting of large, medium, and small game. These data by no means suggest that patterns of subsistence are homogenous over whole continents or subcontinents, but the advocates of subsistence forms based on scavenging or ineffective forms of hunting (Binford 1989; Stiner 1990, 1994) seem to have overstated the case against the existence of reliable hunting economies within MSA and Middle Paleolithic societies (Marean and Kim 1998; Marean and Assefa 1999).

Again in this context, the finds from Schöningen are of central importance and have redefined the discourse on Lower Paleolithic subsistence. Thieme’s (1997, 1999) team recovered eight spears from Schöningen in direct association with the bones of over a dozen horses in deposits dating to ca. 300ka. These discoveries from the mid-1990s brought the more extreme assessment of Lower and Middle Paleolithic subsistence based on obligate scavenging to an end, and as far as I am aware, the implications of these remarkable finds for documenting hunting of large game by archaic hominids and the implications of the recovery of a yew wood spear with the skeleton of an Eemian age forest elephant at Lehringen have not been questioned in recent years. These finds do not demonstrate that hunting large game was a universal phenomenon in the late Middle and Late Pleistocene, but they do document the existence of well-planned and successfully executed hunting of large and fast game.

More mundane sources of information tend to support this view. Numerous faunal assemblages indicate that late archaic and early modern humans had frequent early access to game. In most settings, the possibility of scavenging cannot be completely excluded, but active hunting is the most parsimonious explanation for the faunal assemblages at sites including, for example, Salzgitter-Lebenstedt (Gaudzinski and Roebroeks 2000), Tönchesberg (Conard 1992), and Wallertheim (Schmidtgen and Wagner 1929; Gaudzinski 1995; Conard and Prindiville 2000). In other contexts, in many parts of Eurasia and Africa, similar evidence for the role of mammals in the diet of Middle Paleolithic and MSA people has begun to accumulate (Gaudzinski 1996; Marean and Kim 1998; Marean and Assefa 1999; Burke 2000; Bocherens et al. 1991, 2001). Finally, it must be stressed that scavenging fresh carcasses is an attractive economic option in contemporary hunting and gathering societies (O’Connell et al. 1988). Thus there is no reason to stigmatize Paleolithic scavenging as a premodern adaptation.

In southern Africa, Klein and Parkington have developed new approaches and hypotheses for the development of subsistence practices during the MSA. Parkington (2001) stresses the key role of the exploitation of coastal resources for brain development and the origin of cultural modernity in coastal settings. He has also suggested that similar processes may have driven human evolution in other coastal environments, including the circum-Mediterranean region. Klein
(1999) has looked at small game such as tortoises and marine resources as playing an important role in MSA and LSA subsistence. He argues that until ca. 50 ka, hunting was limited to comparatively easily hunted game and that people only started systematically hunting dangerous animals, including suids and buffalo in the late MSA and LSA. Klein sees this shift in subsistence as an indication of the rise of cultural modernity in connection with genetic mutations and the appearance of fully developed language. Both Parkington’s and Klein’s hypotheses have been received with considerable skepticism, but both hypotheses present entirely welcome, refutable models for the rise of cultural modernity. Given the general lack of clearly formulated models that provide causal explanations for the rise of behavioral modernity, these hypotheses, even if they are later demonstrated to be incorrect, have fostered considerable new research. This is certainly the case of the critical assessment of the early evidence for hunting by Binford and colleagues in the 1980s and 1990s.

Like the other data we have considered thus far, the evidence on subsistence during the Middle and Upper Pleistocene shows a pattern of advanced adaptations at an early date. With the possible exception of Parkington’s model for increased use of marine resources in the Late Pleistocene, the data on subsistence tend to argue against a revolutionary change in economic and social behavior that defines the appearance of cultural modernity.

23.4 Settlement

Reconstructing patterns of settlement and the organization of space is one of the more elusive ways of trying to define modern patterns of behavior. This relates to the general difficulty of reconstructing settlement dynamics in any period and particular problems associated with Paleolithic periods, where the amount and quality of data are generally poorer than in later periods. The analysis of Paleolithic settlement in the contexts of defining modern behavioral forms has two major approaches, one intrasite the other regional.

Binford (1998), Wadley (2001), and others have argued that spatial organization within a find horizon can be used to define cultural modernity. Binford, for example, sees repetitive modular units of hearths and bedding areas in rock-shelters as a hallmark of modern spatial organization. In his view, this pattern of spatial organization is not present before the LSA or Upper Paleolithic. Wadley sees a marked increase in spatial organization during the late MSA of Rose Cottage Cave in the Free State of South Africa as a further indication that the final stages of the MSA may reflect the period in which cultural modernity developed.
In Europe, Kolen (1999) has pointed to the lack of clear evidence for architecture as an indication that neither Lower nor Middle Paleolithic groups regularly built shelters as centers of social and economic interaction, as are known in many later archeological periods. Instead, archaic humans used what Kolen refers to as “nests” to provide shelter. If correct, this would indicate that settlement dynamics of archaic humans, including Neanderthals, fell outside the range of culturally modern people. Several researchers have questioned this model and suggest that even if clear architectural features other than hearths are generally lacking before the Upper Paleolithic, late Middle Paleolithic sites document spatially structured activity areas similar to those one would expect in sites of modern hunters and gatherers (Vaquero et al. 2001, 2004; Conard 2001a). As with many of the criteria considered here, it is unclear to what extent taphonomic factors and the quality of data affect our interpretations. Kolen, however, is certainly correct to note that clear evidence for anthropogenic shelters and dwellings is extremely rare prior to the Upper Paleolithic.

At a larger scale of analysis, we see more tantalizing, yet largely inconclusive evidence for the use of space and distant resources as indicators of behavioral modernity. Important works by Geneste (1988), Roebroeks et al. (1988), Floss (1994), and others examine the use of distant raw materials as a source of information on Paleolithic economic and spatial organization. Especially in the context of the continental European approaches to the study of patterns of lithic reduction and technology (Geneste 1988; Hahn 1988; Boëda et al. 1990), much research has been aimed at linking patterns of lithic technology to systems of mobility and settlement. These and other studies show the nearly universal pattern that more distant raw materials are present at sites in more reduced form than local raw materials. This applies for all Paleolithic periods. In later periods more raw materials from distant sources are transported to sites, but there is no specific moment that reflects a quantum shift from non-modern to modern patterns of behavior. Also, the “provisioning of place” (Kuhn 1995)—that is, the movement of quantities of raw material to sites for future use—is documented on sites of both modern and archaic hominids (Conard and Adler 1997).

Examination of the abundance of distant raw materials as a reflection of the size of territories and long-distance economic and social relationships has also provided ambiguous results. Middle Paleolithic assemblages document the use of raw materials from 100 or more km away (Floss 1994; Fëblot-Augustins 1997). Nonetheless, such long-distance transport of tools and raw materials are still more common in the Upper Paleolithic, and the difference is more one of degree than of kind. So far these kinds of data have not led researchers to devise a reliable means of distinguishing between archaic and modern behavioral forms. These lithic data also suggest mosaic, context-dependent systems of adaptations with
considerable variability, rather than a black-and-white world of unilinear evolution, in which quantum leaps between archaic and modern behavior can be readily identified.

Finally, an analysis of site types and links between sites within settlement systems shows considerable diversity in MSA and Middle Paleolithic systems of settlement, but no easily recognizable criterion for defining behavior modernity (Conard 2001b, 2004b). Here, as in other areas, I doubt whether the search for a holy grail of cultural modernity is a productive way of defining a research program. Scholars continue to struggle to identify the origins of a settlement system that reflects a symbolically mediated landscape inhabited by culturally modern people. Furthermore, if our definition of behavioral modernity includes all ethnographically documented patterns of settlement, we must concede that a nearly endless diversity of adaptations among subrecent hunters and gatherers is by definition modern and by no means easy to distinguish from hypothetical non-modern settlement dynamics as indicated by the distribution of archeologically visible material cultural remains.

23.5 Beyond technology, subsistence, and settlement

As the discussion above suggests, identifying clear criteria for behavioral modernity is probably more likely in the realms of ideology and symbolic communication than in the nuts and bolts archeology of chipped stone and faunal remains. Here, I consider several lines of argument and sets of data that lie outside the pragmatic economic concerns of day-to-day subsistence.

23.5.1 Burials

Most of the more complete human skeletons from before the Middle Paleolithic and Middle Stone Age appear to be the result of extraordinarily favorable taphonomic contexts. Despite arguments to the contrary by Gargett (1989, 1999) and other colleagues, there are a wealth of Middle Paleolithic human skeletons that seem to have been buried deliberately (Solecki 1971; Trinkaus 1983; Defleur 1993). Such burials could be motivated by purely practical factors like the need to dispose of undesirable cadavers, but I think it is more likely that the numerous burials of Neanderthals and anatomically modern humans of the Middle Paleolithic reflect the deliberate burial of kin and are linked to personal and emotional ties between the living and the dead. Defleur (1993) has summarized much of the evidence for Middle Paleolithic burials and points to a number of
convincing cases in Europe and the Levant. The question of the deliberate inclusion of grave goods and the identification of specific ritual practices is more contentious and difficult to demonstrate beyond doubt.

In the Upper Paleolithic the data are unambiguous, and many burials preserve opulent grave goods that reflect the status of the individuals and the needs of the dead in the afterlife. Examples of burials from Sungir', Dolní Věstonice, the Grimaldi Caves, and other sites suggest that the system of beliefs in association with death and the afterlife were much more elaborate in Upper Paleolithic than Middle Paleolithic societies. These Upper Paleolithic burials are universally accepted as indicators of cultural modernity. As far as I am aware, aside from somewhat enigmatic cases like the highly fragmented and partially burnt assemblage from Klasies River Mouth in South Africa, the MSA and early LSA have not produced sufficient data for burials to allow conclusions to be drawn about practices and beliefs in sub-Saharan Africa.

23.5.2 Pigments

In recent years, there have been a number of reports of early occurrences of pigments and discussions of the importance and meaning of the use of pigments (Barham 1998; McBrearty and Brooks 2000; d’Errico and Soressi 2002; Hovers et al. 2003). On the basis of this work, it has become clear that pigments were used in some MSA contexts during the later Middle Pleistocene and in numerous MSA and Middle Paleolithic settings of the Late Pleistocene (Watts 1998). Southern Africa has yielded particularly abundant evidence for the use of pigments during the MSA. Barham’s (1998) work at Twin Rivers in Zambia is a noteworthy example of the presence of many pieces of pigments in Middle Pleistocene contexts, and numerous MSA sites dating to the Late Pleistocene including Klasies River (Singer and Wymer 1982), Diepkloof, Peers Cave, Hollow Rock Shelter (Watts 2002), Apollo 11 (Vogelsang 1998), Blombos (Henshilwood et al. 2001) have produced much evidence for grinding of pigments. Parkington has argued that the use of pigments provides additional indications of the advent of behavioral modernity in the MSA, particularly in more coastal settings, where Howiesons Poort and Still Bay assemblages are concentrated. Watts (1998, 2002) has reviewed the evidence for the use of pigments in the MSA and concludes that they are extremely common at many MSA sites and reflect a widespread ability to structure the world into a symbolically organized whole. Watts rejects the hypothesis that pigments were primarily used for strictly utilitarian purposes, including tanning hides, while Wadley (2005) emphasizes the practical uses of ground ochre, for example, in hafting lithic artifacts.
In the Levant and Europe, Hovers et al. see strong evidence for the use of ochre at Middle Paleolithic sites including Qafzeh (Vandermeersch 1969; Hovers et al. 2003) and Pech de l’Azé (Bordes 1972; d’Errico and Soressi 2002). The potential uses of ground ochre include body painting, rock painting, drawing, ritual, medicinal, as well as more mundane purposes. Although we rarely have reliable information on the specific use of these early occurrences of ochre, they are presumably, at least in some settings, such as in Middle Paleolithic burials, connected with religious beliefs that speak for a high level of cultural development and a significant degree of symbolic communication (Hovers et al. 2003). As with other potential indicators of advanced cultural attributes discussed above, the use of ochre does not appear to reflect a quantum leap signifying the shift from archaic to modern patterns of behavior. Both anatomically modern and archaic humans used pigments and presumably attached symbolic meaning to red, black, and perhaps other ground mineral pigments. Given the well-documented use of mineral pigments, the use of organic pigments is likely, even if difficult to demonstrate with direct archeological observations.

23.5.3 Decorated objects and non-figurative representation

There is a long history of claims for deliberate non-utilitarian modification of objects in Paleolithic contexts. These include finds from the Lower Paleolithic, such as incised bones from Bilzingsleben (Mania 1990; Steguweit 2003), and many finds from later periods. These objects are often controversial and are usually not accepted as demonstrating complex symbolic communication and cultural modernity. Following other lines of argument, colleagues have suggested that the perfect symmetry of some hand axes indicates an advanced aesthetic development, but Wynn (1995) and Haidle (2004) argue that hand axes do not necessarily reflect symbolically based communication or language. Over the course of the Middle Paleolithic and MSA, larger numbers of enigmatic objects have been published, including the cross-incised stone and modified fragment of a mammoth tooth from Tata, Hungary (Vértes 1964), and the so-called “mask” from La Roche-Cotard (Lorblanchet 1999). Some researchers have included evidence for collected fossils or curated natural products as indicators of advanced esthetic and behavioral patterns (Schäfer 1996).

Particularly in recent years, the MSA has produced a number of incised objects that have been taken as evidence for symbolic communication and a high degree of cultural development. Important examples of these finds include engraved linear and cross-hatched patterns on pieces of ochre from Still Bay
deposits at Blombos dating to ca. 75 ka (Henshilwood et al. 2002), and incised pieces of ochre from, for example, Peers Cave. Current excavations at Diepkloof have produced fragments of several decorated ostrich eggshells from Howiesons Poort contexts including a decorated piece of an ostrich eggshell flask (Parkington et al. 2005). Similar finds have also been recovered from MSA contexts at sites including Apollo 11 (Vogelsang 1998). These finds are unquestionably the result of deliberate manufacture and probably reflect the desire of the craftsperson to convey symbolic content and esthetically meaningful information to members of his or her social group. There can be little doubt that such carefully produced decorated objects and the non-figurative representations they carry communicated information from the maker to people who used or saw these objects. Deciphering the specific meaning broadcast through these finds is not easy, and few specific explanations for their meaning have been presented. With increasing amounts of carefully executed fieldwork during the MSA, there is reason for optimism that contextual information will help archeologists to develop hypotheses to explain the meaning of these finds. Some colleagues accept these finds as definitive evidence of cultural modernity with fully developed symbolic communication, modern cognitive abilities including language (Henshilwood et al. 2002; d’Errico et al. 2003), while others are less convinced that these finds demonstrate proof of behavioral modernity.

23.5.4 Ornament

The manufacture and use of ornaments convey social information about individual identity and group affiliation. This means of projecting assertive individual style or emblemic style reflecting social affiliation within a larger demographic group (Wiessner 1983) is an important characteristic of modern behavioral patterns and has been the focus of much recent research (Vanhaeren 2002; Köbl and Conard 2003). The archeological distribution of ornaments provides a clearer signal than many of the classes of finds considered above.

Early evidence for the use of marine shells as ornaments comes from burial contexts from Qafzeh Cave in Israel and dates to about 100 ka (Bar-Yosef and Vandermeersch 1993). Slightly younger examples of perforated marine shell ornaments come from Still Bay deposits at Blombos Cave dating to about 75 ka (Henshilwood et al. 2004). Starting roughly 40 ka, personal ornaments have been documented in many parts of the Old World. Early ornaments include ostrich eggshell beads from early LSA contexts in Enkapune Ya Muto rockshelter, Kenya, with associated radiocarbon measurements between 30 and 40 ka (Ambrose 1998). AMS radiocarbon dates directly on ostrich eggshell beads from deposits
representing the transition from the MSA to LSA at Mumba Cave in Tanzania (Figure 23.5) (Weiß 2000; Conard 2004a) have yielded multiple dates between 29 and 33 ka, and lend support to the early dates from Enkapune Ya Muto. There is every reason to assume that these East African ornaments were made by anatomically and presumably culturally modern people.

Excavations at Ksar Akil in Lebanon (Azoury 1986) and at Üçagizli in the Hatay Province of Turkey (Kuhn et al. 1999, 2001) have produced rich assemblages of perforated marine shells from Initial Upper Paleolithic contexts dating to about 40 ka (Figure 23.6). Similar finds have been recovered from other Mediterranean early Upper Paleolithic contexts, including Riparo Mochi on the Ligurian Coast of Italy (Kuhn and Stiner 1998; Stiner 1999).

Elsewhere in Europe there is considerable evidence for a rapid spread in the use of ornaments with the beginning of the Upper Paleolithic. Neanderthals apparently created a wide range of perforated and incised ornaments in Châtelperronian contexts such as at Grotte du Renne at Arcy-sur-Cure (Leroi-Gourhan and Leroi-Gourhan 1964; d’Errico et al. 1998; Baffier 1999). At more or less the
same time, numerous examples of early Aurignacian ornaments have been recovered from several regions including the Swabian sites such as Vogelherd, Geißenklösterle, and Hohle Fels (Conard 2003a) (Figure 23.7). In addition to incised and perforated natural forms such as teeth, these artifacts include diverse ornaments made of mammoth ivory. It is noteworthy that many of the oldest forms of ornaments in Europe are not only perforated shells or teeth, but also completely carved, three-dimensional ivory beads, pendants, and figurines in which the maker completely dictated the form of the artifact.

Although earlier examples of personal ornament are known, by around 40 ka, examples of ornaments are well documented across much of the Old World. These data are consistent with the hypothesis that modern cultural behavior spread rapidly between roughly 30 and 50 ka. Shell beads from Mandu Mandu Creek Rock Shelter in Western Australia dating to more than 30 ka (Morse 1993) suggest that the use of personal ornaments was indeed widespread at an early date. Although Australia lies outside the scope of this review, the colonization of Sahul was an event in prehistory that required crossing the vast open water of Wallacea with rafts or other forms of boats. The best available dates for the colonization lie in the range of ca. 42–45 ka and fit with the pattern
suggesting the rapid spread of advanced behavioral patterns at about this time (O’Connell and Allen 1998, 2004).

23.5.5 Figurative representations

The presence of figurative art has been universally accepted as an indication of behavioral modernity. As far as I am aware, no one has disputed that figurative representations are a hallmark of modern cultural behavior. Mann (2003) has gone so far as to argue that representational art is the “gold standard” by which behavior modernity can be identified and measured.

In Africa, the earliest figurative art is from the late MSA of Apollo 11, dating between 25,500 and 27,500 radiocarbon years ago (Vogelsang 1998). These examples
of painted mobile art depict a number of animals, geometric forms and a possible therianthrope (Figure 23.8). The Middle Pleistocene aged, anthropomorphic-shaped stone from Tan Tan, Morocco (Bednarik 2003), much like a similar object from Berekhat Ram, Israel (Goren-Inbar 1986; Goren-Inbar and Peltz 1995; d’Errico and Nowell 2000), appears to be a modified natural form rather than deliberately carved figurine. In the Levant there is little or no evidence of figurative art before 30 ka.

The situation in Europe is very different in that several sites have provided evidence of figurative representation between 30 and 40 ka. The earliest figurative art includes the mammoth ivory figurines from four caves in Swabia in southwestern Germany (Hahn 1986; Schmid 1989; Conard and Bolus 2003; Conard 2003b) and several red monochrome paintings from Fumane in northern Italy (Broglio 2002; Broglio and Dalmeri 2005). The Swabian caves of Vogelherd,
Hohlenstein-Stadel, Geißenklosterle, and Hohle Fels have produced about 20, mostly very small, ivory figurines and figurative representations in bone and stone dating well in excess of 30,000 radiocarbon years, which corresponds to closer to 40 ka in calendar years. Due to the noisy radiocarbon signal in this period and above-average $^{14}$C production, the radiocarbon ages at the Swabian Caves and the similarly aged deposits from Fumane significantly underestimate the age of these artworks. The Swabian ivory figurines include depictions of lions, mammoths, horses, bison, bears, a water bird, and two or perhaps three therianthropes that combine features of lions and humans (Hahn 1986; Conard 2003b) (Figures 23.9 and 23.10). These artworks are small and beautifully carved. They stand in sharp stylistic contrast to the highly schematic paintings of animals, unknown forms, and a possible therianthrope from Fumane (Broglio 2002). Geißenklosterle has also produced a painted rock from this period that preserves traces of red, yellow, and black pigments (Hahn 1986).

Most of the spectacular paintings from Grotte Chauvet in the Ardèche region of southern France appear to slightly postdate the examples of figurative art from Swabia and Fumane (Clottes 2001) (Figure 23.11). Here, numerous depictions of animals date back as far as 32,000 radiocarbon years ago. The selection of animals depicted in Chauvet, with an emphasis on dangerous, strong, and large

![Figure 23.9](image-url)

Apollo 11 Cave, Namibia. Figurative painting from Middle Stone Age deposits dated with radiocarbon to ca. 27,000 years old (after Vogelsang 1998)
animals, shows remarkable similarities to the Aurignacian figurines from Swabia and no stylistic similarities to the simple depictions from Fumane. Other important sites in this context include Stratzing in Lower Austria, where a human figurine of stone has been dated to between 30 and 32 ka. Abri Cellier,
La Ferrassie, Abri Blanchard, and Abri Castanet in southwestern France have produced representations of animals and vulvas dating to about 30,000 radiocarbon years ago (Leroi-Gourhan 1995).

These figurative depictions from European contexts are the oldest known worldwide. They all date to the early Upper Paleolithic and were presumably made by modern humans; however, as far as we can tell, Neanderthals still occupied parts of Europe at this time, roughly 40 ka. At present, there is no concrete evidence for a direct association between modern humans and early figurative art in Swabia. Thus, for now, the hypothesis that Neanderthals created the figurative art and other remarkable finds of the early Aurignacian, although improbable, cannot be refuted (Conard et al. 2004a).

The specific context in which figurative art developed has been the subject of considerable discussion of late and will not be elaborated on here (Lewis-Williams 2002; Conard and Bolus 2003). Regardless of the specific social–cultural mechanisms that led to the development and spread of figurative art, there is a consensus among archeologists and paleoanthropologists that the makers of these early artistic traditions were culturally modern people (Churchill and Smith 2001). While many other advanced behavioral forms have precursors in earlier periods, there is no convincing evidence for figurative depictions prior to the beginnings of the European Upper Paleolithic.

### 23.5.6 Music

Perhaps because of the long research tradition and favorable taphonomic conditions, the earliest examples of musical instruments have been recovered from early Aurignacian contexts in Swabia (Hahn and Münzel 1995; d’Errico et al. 2003; Conard et al. 2004b). As is the case with figurative representations, evidence for music and musical instruments can be seen as an indication of fully developed cultural forms based on symbolic communication. The assumption in this context is that where there is figurative art and music, there must have been fully developed language, by which Paleolithic people assigned specific concrete and abstract meaning to words and could efficiently communicate information about the past, present, and future. Thus, where there is figurative art and music, there must have been behaviorally modern people.

While speech, song, music, and dance presumably existed still earlier, the oldest musical instruments known are two bone flutes and one mammoth ivory flute from archeological horizon II at Geißenklösterle (Hahn and Münzel 1995; Conard et al. 2004b). This deposit has been dated by thermoluminescence to about 37,000 BP and to several thousand years younger with radiocarbon.
The better preserved of the bone flutes is made of the radius of a swan (Figures 23.12 and 23.13). Reconstructions of the instrument produce a high-pitched but pleasing music. Friedrich Seeberger (2002, 2004) has recently recorded a CD of Ice Age music played on a reconstructed bone flute of the kind known from Geißenklösterle. This flute can be played without a reed and is clearly a flute rather than a reed- or trumpet-voiced instrument as suggested by d’Errico and colleagues (2003). While Aurignacian musicians may have played very different-sounding music, Seeberger’s playing provides a striking impression of what this early Upper Paleolithic music may have sounded like.

Recent excavations at Vogelherd have also produced fragments of worked bird bones that appear to be remains of a flute, but further work is needed to confirm this preliminary conclusion (Conard and Malina 2006).

Other sites, most notably Isturitz in the French Pyrenees, have produced additional flutes and indicate that wind instruments were in fairly wide use during the early Upper Paleolithic (Buisson 1990; d’Errico et al. 2003). Of course, there are countless other less conspicuous forms of percussion and wind instruments that could have existed during the early Upper Paleolithic or still earlier, yet they remain to be identified. Claims for earlier examples of Middle Paleolithic flutes have generally been met with skepticism in archeological circles, as was the case with recent claims for a Middle Paleolithic flute made from a cave bear bone from Divje Babe in Slovenia (Turk 1997; Albrecht et al. 1998).
23.6 Conclusions

This overview has touched on some, but by no means all, of the evidence for the development of behavioral modernity. I have mentioned some of the main data sets and lines of reasoning that play a role in the discussions and debates about the origins of modern behavior. This leads to the question of by what means, where, and under what circumstances behavioral modernity arose and which of the hypotheses for its origins lies closest to the mark?
The answers to these questions depend on how the evidence is weighed and interpreted. From my point of view there can be no doubt that European Aurignacian societies by roughly 40 ka had all of the hallmarks of modern behavior including Mann’s “gold standard” of figurative art as well as musical instruments. The best evidence for early figurative art and music comes from the caves of the Swabian Jura. While one could argue that some important Upper Paleolithic artifact forms developed in the Upper Danube drainage in the period around the time of the arrival of modern humans, naming this region as the single global center for the origin of cultural modernity would be a radical and naive interpretation. The contemporary finds of figurative art from Fumane, and slightly later finds from southern France, indicate that the beginnings of the Upper Paleolithic reflect a time in which archaic behavioral forms were replaced across the board by behavioral forms that lie within the range of modern variability. This transition appears to have begun across much of Europe about 40 ka when modern humans entered a continent inhabited by Neanderthals. Based on the presence of late Neanderthals in several regions of Europe (Hublin et al. 1995; Smith et al. 1999), it appears that there must have been a period in which both archaic and modern humans coexisted in Europe, and contact between the two forms of people must have occurred. Given the poor chronostratigraphic resolution and lack of human fossil material during this key period between roughly 30 and 40 ka, it is difficult to specify exactly how long both hominids coexisted in specific regions (Conard and Bolus 2003).

Early anatomically modern humans at Skuhl and Qafzeh in the Levant predate many remains of Neanderthals in southwestern Asia and point to an initially successful colonization of the region. The reappearance of Neanderthals in the Levant by roughly 60 ka suggests that Neanderthals had more successful adaptations and demographic advantages over anatomically modern humans in interactions dating to the middle part of the Late Pleistocene. While evidence for strict contemporaneity is still lacking, this observation indicates that in some settings in which both hominins produced Middle Paleolithic artifact assemblages, Neanderthals had the upper hand. However, in later encounters the situation was different. At about 40 ka, modern humans arrived in western Eurasia with better-developed cognitive skills (Lewis-Williams 2002) or behavioral advantages (Marean 2005) that led to demographic success relative to the indigenous Neanderthals. In western Eurasia, a period of dynamic equilibrium between Neanderthals and anatomically modern populations existed, in which moderns presumably profited from the knowledge and cultural practices of the archaics and vice versa.
There is little reason to postulate a violent rapid advance of Neanderthals into the Levant replacing indigenous anatomically modern humans in the middle of the Late Pleistocene, and similarly there is little reason to assume that the arrival and spread of modern humans into Europe was either universally rapid or brutal. On the contrary, the transition from the Middle to the Upper Paleolithic and the infiltration and eventual complete dominance of *Homo sapiens sapiens* in Eurasia probably took on countless local ecologically and historically dictated variants in which there was considerable give and take between archaic and modern humans. This pattern is reflected in the diverse regional signatures of the archeological records from nearly every region that has produced relevant data for this transition. These data show very different archeological signatures depending on the environmental and social–cultural setting encountered by incoming populations (Conard 1998; Conard and Bolus 2003). Evidence from the sites occupied by late Neanderthals indicates that they too manufactured and used ornaments (Baffier 1999), and as we have seen above, there is little that separated the patterns of technology, subsistence, and settlement reflected in Middle Paleolithic artifact assemblages from those of the MSA or early Upper Paleolithic. Still, some time presumably in the early and middle parts of the Late Pleistocene and certainly no later than 40 ka, people began producing material cultural remains that allow us to identify behavioral modernity. This pattern of behavior was carried primarily, but apparently not exclusively, by anatomically modern humans.

Many characteristics of modern behavior can be found across much of the Old World, and the distribution of advanced cultural traits is significantly determined by the intensity of research in different regions. The recent trend of important discoveries being made in MSA contexts in Southern Africa will no doubt continue as more work is done. The data from Klasies River, Apollo 11, Rose Cottage Cave, Blombos, Sibudu, and Diepkloof clearly show the enormous potential of the subcontinent. Elsewhere, a similar intensification of research would perhaps produce a similar increase in data relevant to the definition of cultural modernity. While western Eurasia also has considerable potential, there is less reason to assume that the archeological record will be so radically transformed by further work. Instead, important gaps will be filled and, presumably, gradually a more complete picture of the highly variable behavioral patterns during the Lower, Middle, and Upper Paleolithic will emerge. With time we will be better able to develop and test new hypotheses for the evolution and spread of cultural modernity.

Based on the data presented above, a strict unilinear and monogenetic model for the evolution of behavioral modernity appears less likely than a pattern of highly variable polygenetic development. These data suggest that MSA and
Middle Paleolithic societies generally existed within regionally specialized social groups with highly variable material culture. Whether anatomically modern or archaic, these people lived at a similar level of technological and cultural development. Perhaps by about 80 ka or possibly as few as 40–50 ka, full behavioral modernity developed in Africa and in Eurasia. Most archaic humans appear not to have completely mastered the repertoire of new behaviors including fully developed symbolic communication. If, however, late archaic humans, including Neanderthals, were culturally fully modern, their behavioral patterns still put them at a reproductive and demographic disadvantage in comparison with the anatomically and culturally modern social groups that propagated across the Old World. The extinction of Neanderthals does not necessarily mean that they were not culturally modern, just as the extinction of local groups of Homo sapiens sapiens does not mean that they were not culturally modern. The main characteristic of Homo is that our cultural development can and does vary independent of our biological morphology (Conard 1990). Thus late anatomically archaic peoples may have been behaviorally modern, just as early anatomically modern humans may well have been behaviorally archaic (Zilhão 2001).

In the coming years, archeologists and paleoanthropologists need to establish high-quality regional databases and specific local scenarios and hypotheses for the evolution of modern patterns of behavior (Hublin et al. 1996; Parkington 2001; Lewis-Williams 2002; Conard and Bolus 2003). As work progresses researchers should be able to test these hypotheses and better define these diverse regional scenarios to create new models that come closer to reflecting the evolutionary reality that a nuanced history of our species warrants. This work should proceed using multiple analytical paradigms and shifting scales of analysis (Conard 2001). There are certainly multiple approaches to this complex problem, and all contextually informed explanatory models for the rise of cultural modernity are welcome, regardless of whether they originate from the natural sciences, social science, or humanities.

References

Azoury I (1986) Ksar Akil, Lebanon. A technological and typological analysis of the transitional and early upper Palaeolithic levels of Ksar Akil
Conard NJ (ed) (2001b) Settlement dynamics of the Middle Palaeolithic and Middle Stone Age. Tübingen Publications in Prehistory. Kerns Verlag, Tübingen
Conard NJ (ed) (2004b) Settlement dynamics of the Middle Palaeolithic and Middle Stone Age II. Tübingen Publications in Prehistory. Kerns Verlag, Tübingen
Hahn J (1977) Aurignaciens, das ältere Jungpaläolithikum in Mittel- und Osteuropa. Böhlaub-Verlag, Köln/Wien


Morse K (1993) Shell beads from Mandu Mandu Creek rock-shelter Cape Range peninsula Western Australia dated before 30,000 BP. Antiquity 67: 877–883
Rust A (1950) Die Höhlenfunde von Jabrud (Syrien) Karl Wachholz Verlag, Neumünster
finds from Divje babe I cave site in Slovenia Ljubljana. Opera Instituti Archaeologici Sloveniae 2. Znanstvenoraziskovalni center SAZU, Ljubljana
Vértès L (ed) (1964) Tata. Eine mittelpaläolithische Travertin-Siedlung in Ungarn. Akademie der Wissenschaften, Budapest
Wadley L (2005) Putting Ochre to the test: replication studies of adhesives that may have been used for hafting tools in the Middle Stone Age. J Hum Evol 49: 587–601
24 Paleoanthropology and the Foundation of Ethics: Methodological Remarks on the Problem of Criteriology

Mathias Gutmann · Michael Weingarten

Abstract

From a scientific point of view, human beings are usually referred to within a “criteriological” framework. This raises the seemingly empirical question, of which traits should be considered to be “typically human,” and consequently, the difference between human and nonhuman beings is expressed in terms of an animal–human comparison. Following this type of approach, some methodological shortcomings have to be faced. First, the logical grammar of human activities is reduced to the poor grammar of traits, characters, and features of biological constitution. Second, the argumentative structure of criteriological approaches are reconstructed, by analyzing the premises and presuppositions of fundamental animal–human comparisons. A methodologically sound alternative is approached by referring to a semantically rich starting point in the first step, which exploits the distinguishing mark of human existence, namely human work and human activities. From this mediation-oriented perspective, the nonhuman–human transformation takes place not within (first) nature but within culture. This solution has consequences for empirical research as well as for ethical judgment: ethical approaches lose their putative fundament in empirical descriptions of human beings and paleoanthropology will have to face its often overlooked nondescriptive and normative fundament.

24.1 Introduction

Considering the very fact that paleoanthropology deals with the reconstruction of humanoid and prehumanoid life-forms, one should wonder whether there are
any ethical or at least any normative aspects of paleoanthropology at all. And indeed, even if any related ethical concepts were proved to be invalid, whether based on a general theory of decision, on a presupposed order of preference of values, or on the principles of discoursivity and communicative action, it would not threaten the validity of one single paleoanthropological statement. Hence, at first glance we can assume an asymmetrical relation between ethical reasoning and paleoanthropology regarding at least the validity of its statements about human beings. When we take a closer look we see that, despite this asymmetry, both disciplines have at least one remarkable premise in common. From paleoanthropological as well as from a wide variety of ethical approaches, the subject matter, namely human beings, is referred to as a specific kind of living entity. Even if we were to avoid the identification of “living entities” with “animals,” the underlying definitorial scheme for human beings is a criteriological conceptualization. From a criteriological point of view, human beings may be described as animals with some specific features. This specificity, which applies to the uniqueness of humans, has historically been expressed for example in terms of language ability, tool application, sociability, or religiousness. By definition, the specificity concerned defines humans as animals “plus x,” where “x” stands for the respective character. Accordingly, paleoanthropology is privileged in providing knowledge of the emergence of those characters during the prehistoric phase of human evolution. But a closer look at the possibilities of paleoanthropology itself reveals some intriguing methodological problems, which make this privileged status questionable:

1. If we attempt to identify the characters defining humans, we must refer to a semantic framework that is significantly richer than the resulting description in terms of characters. The priority of a semantically rich description would be true even if we referred to complexes and functionally structured complexes of characters.
2. Even if we consider paleoanthropology to be the main source of knowledge about what makes a life-form human, the transition from a more or less purely descriptive to a normative argument must be justified without reference to this descriptive knowledge itself. This point is of importance when we consider the different kinds of normative structures of the underlying descriptions to structuralize the ethical reasoning in a narrower sense.

According to this difference, we can assume the existence of several normative aspects of description itself without determining their ethical relevance. Whereas normativity is assumed in each relevant aspect of the coordination
and cooperation of human action, ethical reasoning deals in particular with the justification of human actions themselves. Each speech act will show intrinsic normative aspects that regulate the respective and specific expectations and self-obligations connected to the utterance concerned. Thus, promising someone to keep a date is self-obligation on the part of both parties, resulting in mutual expectations. The violation of this obligation, the breaking of the promise, is accordingly not an issue of morality. At least in the first instance, it is a disturbance of the mutuality of expectations underlying the speech act itself. It is at a second remove, when asking the reasons for the violation, that we reach the grounds of ethics.

The justification of human actions (or the justification of an omission) is possible in three fundamentally different ways (following Pfordten 1996):

1. From an anthropocentric point of view, the validity of ethical statement is founded on necessary reference to human beings. This necessity can be understood in many ways—depending on the meaning of the term “human.” Accordingly, a species-specific solution is possible as well as a discourse-ethical or a utilitarian solution. In all these cases the relevant fact is that the authors of an ethical statement belong as elements to the group “human being.”

2. Much weaker than anthropocentrism is an anthroporelational justification. In this case, only some of the steps of ethical argumentation must refer to human affairs, whereas the justification of actions may integrate nonhuman relationships too.

3. The weakest approach (at first glance) is a nonanthroporelational justification. In this case, we would not refer to humans as the main source of moral behavior, but to nonhumane living entities, and sometimes even to nonliving entities, as bearers of values. These values may be moral, esthetic, or economic. They are relevant within the framework of nonanthroporelational justification because they are intrinsic or inherent values, i.e., they are considered to exist essentially, even without any reference to a procedure of evaluation or value ascription. Nonanthroporelational approaches are dominant, e.g., in the field of “eco-ethics” (see the deep-ecology approach of Naess 1989).

In all three cases the most important methodological problem is the definition of human beings as (a more or less relevant) reference group of ethical ascriptions or descriptions. Even in an explicitly non-anthroporelational approach, the specificity of human beings is taken into consideration in the form of a human–animal comparison.
24.2 Ethics and the pitfalls of criteriology

A second taxonomy of ethical approaches is possible that is quite oblique to our first one. Whereas we first dealt with possible authorships and the possible subjects of ascriptions, we will now deal with the procedures that are adopted to reach ethical justification of human actions or omissions. From this perspective, again, a great variety of types of ethical reasoning can be identified, which we summarize here in three fundamental kinds:

1. The results of human actions and their intended as well as their unintended products and consequences is the domain of consequentialist and utilitarian reasoning. This type of justification brings at least three fundamental problems. First, the profit or benefit—described as a function of the respective action—must be identified. In order to do so, an order of relevance or a specific kind of Bentham’s rule needs to be available. Second, the authors of the benefits and losses have to be identified; this can be done by assuming that only those entities that show any form of interests in specific good, or which show a minimal degree of consciousness, need to be taken into consideration. Third, the objects of human actions underlying a specific intervention can be very different in structure. They are assumed to be pure resources and sources of (predominantly) human welfare, and they may even be considered as possible subjects of the ascription of specific rights. These ascensions may be guided by certain criteria, such as the degree of consciousness or possible interests, that the respective targets of the ascriptions are supposed to show or be capable of (this point is of the utmost importance for utilitarian environmental ethics, bioethics or medical ethics; Habermas 1997; Düwell and Steigleder 2003). A prominent example relevant to actual utilitarian (or consequentialist) reasoning is presented by Singer’s discussion of the fundamentals of practical philosophy:

Unsere Isolation ist vorüber. Die Wissenschaft hat uns geholfen, unsere Entwicklungsgeschichte, unsere eigene Beschaffenheit und auch die anderer Lebewesen zu verstehen. Befreit von den Beschränkungen durch die religiöse Konformität, haben wir jetzt eine neue Vorstellung davon, wer wir sind, mit wem wir verwandt sind, wie gering die Unterschiede zwischen uns und anderen Arten sind und von der mehr oder weniger zufälligen Entstehung der Grenze zwischen ‘uns’ und ‘ihnen’. Die Annahme dieser neuen Einsicht wird die Art, wie wir ethische Entscheidungen über Wesen treffen, die zwar leben und zu unserer Spezies gehören, denen aber die Fähigkeiten fehlen, die manche Mitglieder anderer Spezies besitzen, für immer verändern. Warum sollten wir das Leben eines
anenzephalen Kindes als sakrosankt ansehen, uns aber zur Tötung gesunder Pavianse berechtigt fühlen, um ihre Organe zu verwerten? Warum sollten wir Schimpansen infizieren, wenn wir die Vorstellung von Experimenten mit behinderten Menschen verabscheuen, deren geistige Fähigkeiten denen der Schimpansen entsprechen? Die neue Sicht läßt keinen Raum für die üblichen Antworten auf diese Fragen: daß wir Menschen eine besondere Schöpfung seien und allein um unseres Menschseins willen unendlich viel wertvoller als jedes andere Lebewesen. Im Licht unseres neuen Verständnisses von unserer Stellung im Universum werden wir diese herkömmliche Antwort aufgeben und die Grenzen unserer Ethik neu bestimmen müssen. Dieser Revision wird jede Ethik zum Opfer fallen müssen, die auf der Vorstellung beruht, wirklich entscheidend sei, ob ein Wesen ein Mensch ist. Das wird dramatische Auswirkungen nicht nur für unsere Beziehungen zu nichtmenschlichen Tieren haben, sondern für die gesamte traditionelle Ethik der Heiligkeit des Lebens. Denn wenn wir erst einmal von der Annahme abrücken, daß von allen Tieren nur der Mensch eine Art Recht auf Leben hat, dann müssen wir anfangen, nach den Eigenschaften und Fähigkeiten zu fragen, die ein Tier haben muß, um dieses Recht zu besitzen. Wenn wir das tun, werden wir aber, sofern wir dieses Kriterium irgendwo oberhalb des bloßen Lebendigseins ansetzen, die Feststellung nicht umgehen können, daß einige menschliche Wesen dieses Kriterium nicht erfüllen. Dann wird es sehr schwierig werden, weiterhin den Standpunkt zu vertreten, diese Menschen hätten ein Lebensrecht, und dieses Recht gleichzeitig Tieren mit gleichen oder höheren Eigenschaften und Fähigkeiten abzusprechen. (Singer 1998 p 184f)

This citation shows the relevant presuppositions and assumptions: humans are viewed as animals among other animals and in consequence the continuity between biological and cultural, between evolutionary and historical features of human organization, is obvious. From this continuity, some commonly shared features can be derived such as their sensitivity to pain, their consciousness, and their ability to take an interest in specific aspects of their respective environments. But even within the (true) species Homo sapiens, not all members are equipped with these features equally. Consequently for Singer, single members can be equipped with the property of consciousness at least to a certain and individually differing degree, while others are not. And it is this individual difference within the reference group “Homo sapiens” that is relevant for the ethical evaluation of the resulting life-forms that are potentially available for members of the species Homo sapiens and for all nonhuman animals.

2. Principalism and the reminiscence of Kantian thinking. From a very different point of view, principalist reasoning considers a principle necessary, against which a human action may be judged. This principle can be assumed
to be represented by specific inborn characters of the human genus *Homo*. Again, human beings appear to be a token if the type *Homo* is understood as a biological unit:

- Ein guter Grund für die Gattungsbetrachtung ist der Umstand, daß der Ausdruck ‘der Mensch’ das Individuum einer natürlichen Art bezeichnet, deren Mitglieder über Fähigkeiten verfügen, die in der biologischen Artbezeichnung Homo sapiens mit dem qualifizierenden Zusatz sapiens gemeint sind. (Höffe 2002 p 74)

By definition then, humans are reasonable animals, equipped by membership in the genus *Homo* with the capability of rational reasoning. From this point of view, the term “life” remains a biological reference. But nevertheless, some differences can be found within the realm of biological entities, considering the specific type of life respected by the respective natural kinds:

- Die erste Bedeutung, Ethos 1, ist der gewohnte Ort des Lebens, der den Menschen ins Kontinuum der Natur versetzt. Denn auch Tiere haben ein Ethos, das freilich für die jeweilige Art oder Gattung eigentümlich ist. Gemäß ihrer biologischen Ausstattung wohnen Fische im Wasser, das Vieh dagegen steht auf der Weide oder im Stall. Schon bei domestizierten Tieren, eben dem Vieh, tut sich also mehr als eine einzige Möglichkeit auf. Ihr Ethos hängt nicht allein von der biologischen Ausstattung ab, bleibt freilich an diese zurückgebunden. Obwohl durch die Biologie vorgeprägt, ist ihr Ethos durch sie unterbestimmt. (Höffe 2002 p 28f)

The first ethos has animals and humans in common; it refers to the specific place in life, the respective environment of the life-forms. Ethos in the second (tradition) and third (ways of thought, habits, and characters) meanings refers to aspects of life that are exclusive for humane constitution only. The moral attitudes, which are represented by the customs of human communities and the individual characters of humans which are thought to be instantiations of the respective customs and habits, are unique human features.

3. The anthropology of discourse. The central principle of discursive ethical reasoning is a presupposition of full symmetry and reciprocity in participation and access on the one hand, and of obligation on the other hand. Following the concept of an ideal discourse, anybody who might be concerned by a decision should be allowed to participate in the discourse itself. At the same time, full reciprocity and symmetry in accepting the results as well as the consequences of the decision should be given. For the instantiation of a discourse, which is at least oriented at the rules of an ideal discourse, is based on the ability of human beings to participate in rational reasoning, and this ability rests finally upon some fundamental
aspects of human language acquisition. If we play this acquisition back to its evolutionary origins, the scope of ethical reasoning becomes based on the acquisition of some specific human characters, which make all the difference to nonhuman, animal constitution. According to Gehlen’s program of a philosophical anthropology:


From this point of view it is only a small step to assume that rationality and language acquisition are real biological characters, which render the unique human abilities for discourse and discoursivity possible. The human ability to act in accordance with reasons is an inborn character of the true biological species *Homo sapiens*:


And Gerhardt explains the implications of this thesis thus:

also nach seinen eigenen Gründen handeln kann und folglich auch in der Lage ist, sich darüber mit seinesgleichen zu verständigen. (Gerhardt 2001 p 48f)

The orientation of ethical reasoning toward specific characters, which are assumed to be unique for *Homo sapiens* is one of the central features of this kind of approaches and can be found in many other types of ethical approaches, e.g., of a neoaristotelian style or in concepts following Max Scheler or Hans Jonas.

Even when we state that our short taxonomy is neither complete nor treats the topics exhaustively, we recognize a remarkable similarity in what is presented: all refer (at least in the form of a “weak naturalism,” see Habermas 1999) to a criteriological description of human entities. From this point of view, the transformation from nonhuman to human entities takes place within the realm of nature. Humans then are and remain “natural entities” and the difference from animals should be definable in terms of “new” evolutionary characters (autapomorphies in a strict sense of the word). This similarity brings us back to the central assumption of our chapter: by applying characters within an animal–human comparison, we run the risk of homonymy if we do not carefully examine our theoretical frameworks. The terms “language” or “action” or even “cognition” may belong to much richer logical grammars when we deal with them in a purely descriptive biological framework in contrast to their application in an evaluative, normative, or even ethical framework. To make this assumption plausible, we will look for the problem of characters in a clear context, namely that of evolutionary theory.

### 24.3 The problem of character

Dealing with human beings as animals, we are confronted not so much with normative but with methodological problems concerning the concepts of “characters” which are applied in evolutionary theory. From a systematic point of view, the situation seems to be clear-cut. A great variety of living entities, of animals and plants, exist, presenting an overwhelming magnitude of characters. Accordingly, the most important task of the systematicist may be seen in the identification of those characters that are useful or necessary for evolutionary reconstructions. Following this approach, we need to discern between characters that are more or less useful. But on closer examination, this simple picture becomes somewhat confused:

1. We can establish the difference between “useful” and “less useful” characters only if a criterion is already available, one that permits determining given
characters. If the justification of such a criterion refers to evolutionary knowledge, we run the risk of a vicious circle or—alternatively—an infinite regress (Peters and Gutmann 1971, 1972; Weingarten 1992). This is true on the level of the characters compared, as well as on the level of principles that are applied to evaluate resulting trees:

(a) If we fix the difference between plesiomorphic and apomorphic characters on the one hand, and homologous and analogous characters on the other hand, the methodological problem becomes immediately clearer. But:

(I) In the first case the direction of evolutionary transformation must be known.
(II) In the second case at least the membership of the animals under consideration in the narrower systematic group must be known.

Hence, in both cases evolutionary knowledge must be invested to some extent. If such knowledge is justified based on the results of its application, a vicious circle results. If it is justified with reference to further evolutionary knowledge, an infinite regress results.

(b) If we apply optimization principles, such as parsimony or minimum evolution, we will not avoid this problem because it is methodological and not empirical. For the justification of these principles we can adopt two strategies:

(I) On the one hand we might assume that, e.g., “parsimony” is a “natural” principle:

* The first is that parsimony is a methodological convention that compels us to maximise the amount of evolutionary similarity that we can explain as homologous similarity, that is, we want to maximise the similarity that we can attribute to common ancestry. Any character which does not fit a given tree requires us to postulate that the similarity between two sequences shown by that character arose independently in the two sequences – the similarity is due to homoplasy not homology. (Page and Holmes 1998 p 190)

1 The last way out of this dilemma could be seen in a “hermeneutic” interpretation of the given characters “in the light” of some preceding knowledge. This solution fails unless the criteria for the selection of valid evolutionary knowledge itself is provided. So on a higher level of argument, we finally reproduce the original problem. (For the possibilities of an “evolutionary hermeneutics,” see Gutmann & Hertler in prep.)
An ontological “solution” of this kind can be rephrased as a statement on the “nature of evolution.” In this case, we would have to justify the knowledge about the nature of evolution we attempted to elaborate by the application of the principle. So either the principle is not necessary for the task of evolutionary reconstruction (because we already know that evolution follows this principle), or the justification leads to another vicious circle.

(II) Alternatively, we can assume that parsimony is simply the result of methodological decision.

> The second view is that parsimony is based on an implicit assumption about evolution, namely that evolutionary change implies that the tree that minimises change is likely to be the best estimate of the actual phylogeny. Under this view, parsimony may be viewed as an approximation to maximum likelihood methods (…), and indeed it was in this context that parsimony methods were first proposed by Edwards and Cavalli-Sforza. (Page and Holmes 1998 p 191)

If we alternatively introduce the principle as a convention, and hope that it proves fruitful in application, the shortcomings of the naturalized alternative are avoided. Hence, at the same time as we fail in our purpose of justification, because the convention itself remains unsettled the principle of parsimony must be regarded to have a nonbiological origin. If we alternatively consider, that parsimony refers not to the “natural process” itself, but to the level of hypothesizing this process, the methodological problem of the criteria for the evaluation of the hypothesis reappears:

> Die sparsamste Erklärung für die Deutung von Identitäten der Merkmale von Organismen beruht auf der Schätzung der Wahrscheinlichkeit, daß Übereinstimmung auf eine gemeinsame Ursache zurückzuführen ist: Es ist wahrscheinlicher, daß eine komplexe Kette von Ereignissen (die Evolution eines kompliziert aufgebauten Organs) nur einmal stattfand, als daß zufällig dieselbe Folge mehrfach auftrat (…). (Wägele 2000 p 41f)

In this case, the “likelihood” or the “probability” of the hypothesis is tightly connected to a more general assumption on the natural process, e.g., that the unique development of “complex” organs appears to be more likely than the multiple generation of such characters. But here again, the problem is one of language. The term “complex” is a “polar-contrary-term” and as such is a multivalent predicate. From this point of view the phrase “A is complex” is simply incomplete and ill-formed, and should be replaced by “A is more complex than B in reference to a standard S.” And here again the methodological problem arises of how to justify the standard. One “solution” may be seen in a classical “authority-approach,” defining homologies as those structures that are identified
as good characters by a skilled and experienced expert. The criterion for his skillfulness is of course precisely his ability to recognize good characters, and thus the circle becomes closed. Alternatively, we could attempt to assess the likelihood of a given tree. This likelihood must not be confused with the “correctness” of the tree itself; it is simply an immanent criterion for the validity of a tree in reference to a given data set:

- It is important to distinguish between the probability of getting the observed data, and the probability of the underlying model being correct. Likelihood says nothing about the probability of the model itself. (Page and Holmes 1998 p 155)

If we confuse the criterion of model quality with an assessment of the “reality content” of the model, we commit a simple mistake of category (for further reading see Sober 1988; Page and Holmes 1998).

2. We must consider that the features we use for taxonomic goals—i.e., the construction of a sound “order” of animals or plants referring to a set of these features—are not only of a practical value but must also be thought to express systematic relations. This leads to the assumption that the features we picked up for taxonomic purposes serve equally well in different, nontaxonomic contexts, e.g., for the purpose of physiological, genetic, ecological, or evolutionary research. From a methodological point of view, this “naturalistic” interpretation of scientific description implies that living entities “consist” of characters that they simply “have.” Accordingly, we may assume that the identification of characters which are, e.g., applied for the construction of hierarchies, are based on “innate” skills of the observer himself:


This variation of the well-known Spencerian concept of “evolutionary theory of knowledge” has to face the fact that even if we connect the ability to classify with some innate skills, this does not imply the correctness of the classification (for similar approaches see Oeser 1987, Vollmer 1987; for a methodological refutation see Janich 1987). Consequently, the essential problem of the criteria that allowed the evaluation of the adequacy of the resulting classifications means that the procedures of classification cannot be justified by referring to “innate abilities”
again. Otherwise, the absurd consequence will follow that the correctness of formal procedures, such as the formation of sets in mathematical theory, depends on the “correct” innate skills on part of the mathematician, i.e., the correctness of set-theoretical construction would depend finally on the correctness of biological descriptions (i.e., those descriptions that allowed pursuit of the correct abilities). And it is exactly this biological knowledge that has to be proven as valid knowledge. Consequently, the reference to biological knowledge for the sake of theory of knowledge leads to a simple fallacy of category. In order to avoid such strong assumptions, we should assume that the identification of characters depends directly or indirectly upon the context of the respective description. The reference to the context of description will have at least two consequences:

(a) It is this context that has to be considered if the taxonomic description of living animals is to be transcended. In the case of systematic relations between living animals, we will have to refer to different theoretical contexts, as in the case of genetic or ecological relations. From this point of view, the conclusion of Wagner and Laubichler is inevitable:

- The objects and functional characters that are at the centre of every biological theory are not just found objects waiting to be collected in nature, rather they are conceptual abstractions determined by the parameters of a specific theoretical interest (…). As such, they can only be identified within the context of the object that is ontologically prior to them. Therefore, a precise formulation of these objects of ontological primacy has to be part of every theory as well. (Wagner and Laubichler 2001 p 145)

We will then see whether the solution of the problem of character identification necessarily depends on ontological assumptions, or whether we can give a methodological alternative instead.

(b) If the identification of characters depends on the specific (theoretical) context, a language of description must be applied that allows a distinct and univocal use of designations. Thus the term “gill” is applied to salmo and to pacifastacus in a homonymic, not in a synonymic, way. In this case, we must refer to a framework that allows the localization and comparative evaluation of characters. So we must arrive at theory of organism even before introducing such elementary evolutionary concepts as the “Bauplan” of a living entity. Wagner and Laubichler express this insight in terms of ontological considerations:

- In other words, we assume the ontological primacy of organisms and derive the objects relevant to the theory, i.e., the biological characters, by means of a
conceptual decomposition of the organism. We further argue that if we define biological characters as conceptual abstractions, we also need an appropriately defined organism concept within biological theory. (Wagner and Laubichler 2001 p 144)

Both results lead to the conclusion that one of the main problems of biological descriptions is the problem of comparability. Only if the parts of two animals are comparable in an explicable sense (e.g., systematically), is the application of the same (e.g., systematic) term justified. According to this criteriology, the definition of humans usually refers to a comparison with animals. Even if we consider this problem to be solvable in principle, we would have to face another interesting task in order to assume that humanoid organization is a continuation of animal constitution: how would we have to deal with those human characters for which we cannot find any antecedent aspects within the realm of animals? In order to gain a starting point for tackling the problem, we must reconsider the specific situation we find by dealing with extinct predecessors of humanoid organization.

### 24.4 From remnants to organisms

But even if we tried to apply a character-based description in order to compare humans with nonhuman entities, a specificity becomes observable in each kind of paleontological research, a specificity which is relevant in biology when we reconstruct the evolutionary pathways of nonextinct species. The simple fact is that when we deal with extinct species we deal with the remnants of living entities. This situation increases the difficulties that we pointed out for criteriological approaches:

1. The transformation of the respective organism has to be expressed in terms of the transformation of individual characters or character complexes. Consequently, even functional or reproductive relations must be expressed in terms of characters (Wood and Collard 1999; Wagner and Laubichler 2001).
2. We need to assume a representational relation between the organism and those characters used for the evolutionary reconstruction. This leads either to the supposition that character-states of current organisms represent evolutionarily older character-states (e.g., by maintaining older character-states), or that current organisms themselves represent an ancestral constitution, “living fossils.”

   In both cases, the line of evolutionary transformation has to be already available for the identification of relevant characters. Tattersall (1997) underlines
the simple fact that it may be impossible to express reproductive continuity in terms of character sequences. The tension between the description of human predecessors and the requirements of a population genetic description of reproductive structure seems to be an unbridgeable gap resulting from an empirically insufficient situation:

1. Even if all prehuman fossils found as yet are taken into account, the empirical basis of a sound variation analysis seems to be rather speculative. Without such an analysis species identification is restricted to a rather typological differentiation.
2. The reproductive patterns of prehuman populations are assessed in terms of character patterns rooted in typological ordering itself.
3. Fossil remains as well as representatives of genetic traits need to be understood as representatives of specific cultural abilities and technological achievements.

This inconvenient situation is not just the result of a (undoubtedly insufficient) database; it is a reflex of the specific situation of paleontology as a scientific project. Because “characters” need to be integrated into a framework within which they may serve as (biologically) meaningful features of living entities, we have to interpret the remnants we find in terms of fossils in the light of our biological knowledge. This “interpretation” leads in the first instance not to character sets but to a functional design, to body-plans or constructions. These constructions do not represent living organisms, but they represent “possible living organisms,” which may serve as the predecessors of the living entities that built our starting point. Those functional descriptions, which finally lead to a functional plan, refer to functional knowledge borrowed from nonbiological sources (for the procedure, see Gould 1970; Rudwick 1998; Gutmann 2002a; Morris 2003). In describing animals (of nonextinct species) as if they were functional units, we gain the very starting point of our reconstruction of possible predecessors. One should bear in mind that these constructions are structuralizations of living animals. Thus the transformation next described must be considered as a transformation sequence based on biomechanical principles. These principles can be borrowed from engineering practice. Describing animals as if they were functional units (or “constructions,” see Gutmann 2002a), the transformation of these construction can be described by either optimizing them or by differentiation of a given construction for different working conditions (for the logic of “as if statements” see Kant 1988 and Lewontin 2000). But because animals are only treated as if they were mechanical constructions, some fundamental differences to the engineering of machines or engines must be realized (Gutmann 2005):
1. In contrast to the optimization or differentiation of engines, we may start the evolutionary reconstruction with the construction of actual existing forms and go back to those constructions that can be regarded as predecessors. We can call this procedure “retroengineering.”

We can formally summarize this procedure as follows:

\[
K_n(A_{K_{n-1}}^{RP}) \Rightarrow \ldots \Rightarrow K_4(A_{K_3}/P_{K_3}) \Rightarrow K_3(A_{K_2}/P_{K_2}) \Rightarrow K_2(A_{K_1}/P_{K_1}) \Rightarrow K_1^E(P_{K_1})
\]

Formula 1: \( K \) = the construction of an animal, with “\( E \)” indicating the terminal construction, i.e., the currently structuralized animals. \( A \) = the “antece- dent” and “\( P \)” the postsequent” of a construction. \( RP \) = the “reconstruction principles.”

2. In contrast to the optimization or differentiation of engines, all the single transformational intermediates in building a transformation line must be regarded as being “fit for work.” In terms of engineering processes, the force, form, and material closure of the intermediates of a transformation line must be maintained during the transformation process. Additionally, the construction of a field of this type provides the possibility to reconstruct more than one predecessor for one given starting construction. We can formally summarize this procedure as follows:

\[
\begin{align*}
&K_2b \\
&\uparrow \\
&K_{3_a}^{RP} \Rightarrow K_{2_a}^{RP} \Rightarrow K_{1}^{E} \\
&\downarrow \\
&K_{2_c} \\
&\uparrow \\
&K_4^{RP} \Rightarrow K_{3_b}^{RP} \Rightarrow K_{2_d}^{RP} \Rightarrow K_{1}^{E} \\
&\downarrow \\
&K_{3_c}^{RP} \Rightarrow K_{2_e}^{RP} \Rightarrow K_{1}^{E} \\
&\text{DF}_3 \quad \text{DF}_2 \quad \text{DF}_1
\end{align*}
\]

\( \{ \text{TR}_1 \} \quad \{ \text{TR}_2 \} \quad \{ \text{TR}_3 \} \)
Formula 2: “$K$” = the construction with $E$ indicating the terminal construction of actually structuralized animals. RP = the “reconstruction principles,” DF = the fields of differentiation, and TR = the lines of transformation.

3. The antecedent constructions of a given construction must be “reinterpreted” as the biomechanical “conditions of possibility” of (formerly) living entities. This procedure—i.e., the “imagination” of living entities on bases of the constructional description within a transformation line—can be called “reverse engineering.”

Formula 3: The legend is the same as in 3. Additionally $Lw = “reinterpreted” animals, referring to the respective constructions.

\[
K_n(A_{K_{n-1}})^{\text{RP}} \Rightarrow \ldots K_4(A_{K_3}/P_{K_3})^{\text{RP}} \Rightarrow K_3(A_{K_2}/P_{K_4})^{\text{RP}} \Rightarrow K_2(A_{K_1}/P_{K_3})^{\text{RP}} \Rightarrow K_1^E(P_{K_3})
\]

For the elaboration of a reconstruction, the entire construction has to be regarded as a coherent mechanical unit. Consequently, we cannot reduce the reconstruction to the transformation of some features or characters. In contrast, we can identify the features that may serve as “characters” of evolutionary transformation only after the reconstruction is done. The methodology of an “evolutionary hermeneutics” (see Reif 2004; Gutmann 2002a) urges us to anticipate a comprehensive description of those animals that serve as a source of knowledge, on the functional constraint that we may be faced with when dealing with fossil or geological descriptions of ecological settings within which an extinct lifeworld has to be placed. Asking to characterize human origins now, hoping for some key characters which once achieved serve to explain the transformation from animal to nonanimal constitution, we are confronted with the same problems.

### 24.5 Human characters

When we treat humans as objects of anthropology and replace the term human by “*Homo sp.,” the grammatical structure of our designation implies that humans are animals of a certain kind. Following a common strategy of biological research, we identify characteristics of those animals that are supposed to be human. By conceptualizing humans as animals we expect to locate their origin within the
field of animal organization. We should thus be able to apply those tools we frequently use to identify the origins of nonhumans.

The assumption that human beings are entities with specified properties or characters to be determined by comparison with other entities, such as animals or plants, forms the rational core of criteriological concepts. From a methodological point of view, anthropology becomes a general ontology of being human. Because of human beings’ ontological eccentricity, their “Sonderstellung” within the realm of living entities, anthropology is usually accompanied by cosmological considerations:

> In the first mythological explanations of the universe, we always find a primitive anthropology side by side with a primitive cosmology. The question of the origin of the world is inextricably interwoven with the question of the origin of man. (Cassirer 1972 p 3)

Both discourses, anthropological and cosmological alike, aim at the same question, i.e., the problem of origin. Human nature is rooted within nature itself, i.e., among the variety of other, consequently nonhuman beings or entities. It furnishes the genus Homo with specific and unique properties, abilities, etc. The embedding of human nature into Nature as a whole is reflected by the underlying substance ontology, because it provides the general framework within which all the different entities, constituted as substances, will find their respective places. Relevant aspects of the human constitution are usually expressed in terms of defining characters such as “l’homme machine, Homo faber, Homo pecans et redivivus, Homo sapiens, Homo mechanicus, Homo sensitivus, Homo negans et patiens, Homo patheticus et divinans, Homo curans, Homo vitalis (Homo oeconomicus, Homo politicus, Homo libidinosus) super-Homo, Homo geometricus” (Scheler 1988 p 22f).

According to this argument, we start with a comparison between animals and humans. A comparison constitutes a three-termed relation between both of the things compared. From a logical point of view, we are dealing with at least two language levels, which we may respectively call the level of object-language and the level of metalanguage. Determining the use of terms in metalanguage, we establish the identity of two expressions in object-language. Irrespective of differences, we eventually realize that we are dealing with the two expressions as if they were identical. Consider, for example, the four expressions: “black,” “schwarz,” “noir,” and “niger.” Independent of their application as predicators in all relevant contexts of speech, the four terms are exchangeable with reference to their meaning. In this case we call them synonymous. Referring to their synonymy, the respective expressions are intentionally identical, which implies analytically their extensional identity. According to the number of the letters
used, “black,” “niger,” and “white” are similar, unlike e.g., “noir” and “schwarz.” When we use an abstractor (here: synonymy), we neglect differences in a standardized language (metalanguage), which occur within object-language (i.e., the respective common sense use of language). Simultaneously, metalanguage extends the possibilities of dealing with those expressions not available within object-language. To compare humans with animals, we thus need first to determine the criteria of identity expressed in terms of sameness. Animals and humans share, e.g., the same metabolism, or the DNA sequences of humans and chimpanzees are 98% identical. Referring to sameness for abstraction, we do not in fact consider humans to be animals. We simply treat them like animals in certain respects. From this point of view, we may succeed with the characterization of humans as animals by applying one of two alternative strategies.

Both characters and features of humans are conceptualized homologically as characters and features of animals. All defining characters—even those which are newly acquired with respect to the last common ancestor—are derived from precursors. This inference implies one of the following substrategies:

1. Human characters are directly derived from their precursors. Even characteristic human properties, such as culture, are understood to have counterparts among animals.
2. Human characters may be acquired by emergence, i.e., without immediate precursors, or the characters and features of humans are understood as representing features of animals in an analogical sense.

All three (sub-) strategies may be exemplified by empirical studies. We will restrict ourselves here to their respective methodological consequences. Following the first line of argument, humans are animals. Consequently, humans are organisms and, as organisms, character-bearing systems. Comparing organisms with each other allows us to trace the history of human-specific characters:

► Thus, scientific inquiry based on a homological line of reasoning always has ready justifications. Reconstructions of our phylogenetic past serve the understanding of our morphological design, as well as psychological and behavioural dispositions. Comparative analysis elucidates which features we share with other species and which combinations of traits are different, in particular with respect to ecologically similar species. Information like this can illuminate biological foundations (i.e., constraints and conditions, of human culture). (Weingart et al. 1997 p 151)

In this case, human nature becomes fundamental for culture—itself a specific character of humans. Culture then is by definition acquired during evolutionary
transformation, i.e., compared with the plesiomorphic features found exclusively among nonhuman animals. The methods for the reconstruction of evolution may thus also be applied for the reconstruction of its development. The development of culture is understood as an evolutionary transformation from basic features toward the specific expression of this character, its apomorphic state being found exclusively among humans. Accordingly, the transformation of cultural features needs to be explained in terms of evolution. The difference between nature and culture is embedded within nature, namely as the nature of humans as animals. The descriptions of human nature as the subject matter of evolutionary biology result in the determination of features invariant of culture shared by all members of the natural kind Homo. These features then provide the universals assumed to exist independently of cultural differences. Additionally, culture itself is a feature underlying evolutionary forces or mechanisms. This addition in particular characterizes the homology approach.

Consequently, we need to explain culture as a continuation of animal properties. We should at least be able to relate every single aspect of culture to an underlying feature, performance, or character of nonhuman predecessors. The absurdity of such an attempt becomes immediately evident when we try to determine the fitness value of one of Beethoven’s symphonies. Setting aside, for the sake of the argument, the empirical problems that accompany such homology approaches (discussed by Lewontin 1961; Lewontin et al. 1984; Gould 1988), a wide array of methodological problems remains to be solved. The methodological shortcomings are closely linked with the applied language, and result from a confusion of different levels of descriptions. The explanandum of a homology approach is determined on a language level, absolutely free of any biological aspects. In the case of our musical example, we may in fact describe the harmonic and structural development of a classical symphony without referring to any biological or evolutionary implications. With reference to evolutionary theory or population genetics the term “music” is simply meaningless.

Alternatively, we may attempt to explain the biological role of music. In this case, not music itself but its biological role provides the explanandum. We thereby reproduce the methodological shortcoming of our previous explanatory attempt. Now we need to derive music from its biologically described fore-runners, which implies to determine the fitness values of classical music per se (Eibl-Eibesfeldt 1997).

A weaker assumption is provided by analogy approaches. Following this line of argument, humans are not treated as animals but like animals. Analogy approaches claim that the cultural unfolding of humans resembles the evolutionary process considered to govern the development of animals in certain restricted aspects. Considering this decisive methodological difference, the central problem
of the homology approach, i.e., the necessity to claim the same mechanism for cultural and for evolutionary change, is avoided:

- The basic differences between changing cultures and changing biological populations preclude mere subsumption. It is commonly understood that new cultural practices can be produced in direct response to perceived needs, unlike the randomly generated novelty in evolutionary processes. It is sometimes supposed that this process makes sociocultural theories in terms of the Darwinian model impossible. (Weingart et al. 1997 p 286)

Cultural development is released from the biological basis ruled by evolution. The mechanisms underlying both of the processes may differ. This again results in a dualism between culture and nature. This dualism is, however, compatible with the concept of analogy applied here. The formal structure of an analogy is similar to that of a comparison, i.e., one thing or process resembles another thing or process with reference to a particular criterion. If A and B have the property Q and A has additionally the property P, it may then be inferred by analogy that B has the property P as well. Obviously, this conclusion is only valid if Q implies P. Analogical relations are of some interest for empirical research and usually applied for heuristic purposes, in particular if the two relata come from different genera.

An excellent example is provided by comparing the principles of aerodynamics of an airplane and bird flight. It can be stated that both of them, airplane and bird, are capable of flying. One of the necessary conditions for takeoff is the production of sufficient buoyancy, and this may be verified in the case of bird flight. On the other hand, the identity of the mechanisms for the production of the necessary forces cannot be inferred. This corresponds to the use of the term analogy, when humans are considered to be like animals. Let culture be the explanandum, and let the explanatory relation between species and the processes by which they originate be one part of analogy. An analogy can only be drawn between the evolution of species and the development of culture, if this analogy of processes is determined in terms of reproduction. In this case, the analogy may lead to a reconstruction of the “phylogeny of culture” (Boyd et al. 1997: 364). The extent of similarity between species and cultures may be described within a range of correspondences. The analogy between species and cultures transfers those aspects of species that are necessary prerequisites for phylogenetic analysis, reciprocal reproductive isolation, hierarchically integrated wholeness or individuality. By claiming this similarity, the analogy argument leads to the identical treatment of species and cultures in terms of phylogeny. In order to reconstruct the phylogeny of culture, we need to identify characters adequate for evolutionary analysis. We need to identify homological characters in contrast to purely
analogical ones. In this case, the main argument against phylogenetic methods can also be applied in the context of the reconstruction of “cultural evolution.” When we distinguish between homological and analogical characters in order to choose those characters adequate for an evolutionary reconstruction, the course of evolution needs to be known beforehand. The reconstruction itself becomes a truism. Reconstructing phylogeny on the basis of those homologies leads inevitably into a vicious circle (for further reading concerning these methodological shortcomings of phylogenetic method, see Peters and Gutmann 1971, 1972; Weingarten 1992; Gutmann and Janich 1998).

Besides all this, a central problem concerning the formal structure of the analogy argument remains. Let species and cultures have the characteristics required for phylogenetic reconstruction, for example, being isolated, integrated wholes; and let species show evolutionary processes of a distinct type. The conclusion that cultures show evolutionary processes of this type is only true if the shared characters imply that those units show evolutionary processes of the respective kind. This assumption cannot be justified by referring to analogy again. Alternatively, we might try to justify this analogy by stating that culture-producing humans are animals and not just like animals. In this case, the analogy argument collapses into the homology argument, and similar objections become relevant.

Irrespective of differences in detail, both approaches lead to the same unsatisfactory conclusion, namely that the treatment of culture on the basis of a character-oriented approach arises from a fallacy of category—indicated here by the confusion of two language levels. We may gain an alternative by defining culture not in terms of characters but with reference to our definition of a methodological starting point as the expression of human activity.

### 24.6 Culture as a result of human work

As we saw, humans can be defined in terms of culture. The term “culture” comprises all the differences we may identify in comparing humans with animals. However, we just demonstrated that culture cannot be treated as a character without running into logical and methodological difficulties. Mediation-oriented approaches provide an alternative. Both aspects, nature and culture, are interpreted as the constituents of a relation, not as specifically given determinations as such. Cassirer’s functionalist approach of symbolic forms exemplifies this argument. According to Cassirer’s approach, the relation between nature and culture cannot be defined in terms of derivation. Neither the reduction of human existence to natural states nor the exclusion of natural aspects from the human world is required, but the reconstruction of the relation between culture
and nature. This relation is elucidated to its interindividual and intersubjective roots:

- Truth is by nature the offspring of dialectic thought. It cannot be gained, therefore, except through a constant cooperation of the subjects in mutual interrogation and reply. It is not therefore like an empirical object; it must be understood as the outgrowth of a social act. Here we have the new, indirect answer to the question “What is man?” Man is declared to be that creature who is constantly in search of himself—a creature who in every moment of his existence must examine and scrutinize the conditions of his existence. In this scrutiny, in this critical attitude toward human life, consists the real value of human life. “A life which is unexamined”, says Sokrates in his Apology, “is not worth living.” (Cassirer 1972 p 5f)

The specific human reference to the world structures the search for us. This reference is a form of mediation allowing us to preserve the specificity of nature as well as culture, and to recognize their inner mutual connection. Nature and culture correspond to each other in terms of a contrast, not of a simple continuity. This mutual reflection (sensu verbis) refers to a common medium in which natural and cultural aspects of human activity are constituted. The medium is constituted by “symbolic forms,” for example, language, myth, art, knowledge, etc. It enables, structures, determines, and restricts human reference to the world. In contrast to the attempts described above, this approach refers primarily to all the objectivations of human activity. They represent a collective medium beyond individuality in regard to which relations between two subjects, as well as between subject and object, are constituted. The very structure of this mediation is not restricted to a simple dyadic relation. The polyadic relation (i.e., the nature–culture difference within a symbolic form, here: the form of scientific knowledge) allows us to describe the human world as a region of mediated immediacy. In this connection, the constitution of human beings as cultural entities may be reconstructed. Consequently, humans cannot be conceptualized as a simple transition from nature to culture, but as an articulation of differences between nature and culture during the cultural development of the forms themselves. Nature is not the other side of culture; both terms refer to the same referent, i.e., to human activity, more precisely to human work:

- Man’s outstanding characteristic, his distinguishing mark, is not his metaphysical or physical nature—but his work. It is his work, it is the system of human activities, which defines and determines the circle of “humanity”. Language, myth, religion, art, science, history are the constituents, the various sectors of this circle. (Cassirer 1972 p 68)
Work is understood as collective activity mediated via tools. In work, the difference between nature and culture is constantly produced and reproduced. Reproduction in this sense does not refer to genetic reproduction but to the development of tools, the application of which generates the difference between nature and culture. This activity indicates the starting point for a symbolic reconstruction. Human work can neither be reduced to simple organic activity nor equated with production. It is characterized by the procedure, the means and tools applied, the producers, and the resulting products. Consequently, relations between product and procedures are taken into account, not only as the process of the correct choice of adequate means to reach a certain goal. The development of the relation between means and goals, means and means, and symbols and symbols provides a comprehensive definition of humanity. Such a generic definition comprises two aspects of the term development. The systematic relations between the constituents, and the historic relations as their respective objectivations, are taken into account. The “given-ness” of symbolic forms, their historic expression, and systematic unfolding represent aspects of this definition.

This answers the question of the meaning of the explanandum, insofar as a developmental point of view provides a basis for criteriological characterization. Humans may be defined by their mutual relationship within their society on the one hand and to their environment on the other. In this context, the term “environment” does not serve as a biological expression. Referring to human work, it comprises those aspects of the surrounding of human societies and communities, defined e.g., as resources of the production processes, the entirety of applied tools and means.

According to Cassirer’s concepts of symbolic constitution and the reconstruction of human activity, all these activities are mediated by the respective means of production. If we want to understand the environmental relationship of humans and their communities we have to refer to these means of production, which are the tools of social reproduction. The same holds true if we aim to reconstruct the transformation of humans. The transformation of human activities themselves represents the starting point for the understanding of transformation of human beings. Finally, it should not go by unnoticed that Cassirer’s considerations reflect the limitations of a classical background. Symbolic forms are derived from the field of more or less canonical and traditional determinations of human objectivation. Nevertheless, the concept of the symbolic serves as a basis for a generic definition of humans as developing entities leading toward the wide and seemingly unrestricted area of functional characterization of humans beyond the narrow limits of substantialism:
Language, art, myth, religion are no isolated, random creations. They are held together by a common bond. But this bond is not a vinculum substantiale, as it was conceived and described in scholastic thought; it is rather a vinculum functionale. It is the basic function of speech, of myth, of art, of religion that we must seek far behind their innumerable shapes and utterances, and that in the last analysis we must attempt to trace back to a common origin. (Cassirer 1972 p 68)

The double process of analyzing the expressions of mediated human activity and tracing their origins represents a central issue in mediation-oriented approaches. We are now prepared to address the first of the three questions raised in the introduction. In order to elucidate what is meant by explaining human origins, we will apply our results to the analysis of defining humans in terms of mediation-oriented approaches.

24.7 Defining humans by their activities: the problem of the methodological starting point

Reconstructions of the kind presented above are adequate tools for historical explanation. Such explanations presuppose a semantically rich description as a methodological starting point. The richness of this description contrasts with less rich descriptions, e.g., scientific descriptions because the philosophical grammar of the human world differs fundamentally from the grammar of physical or biological descriptions. Kambartel (1989) identified a fallacy of category here if we fail to recognize the reduction of semantic richness. He demonstrates that methodological problems exist with evolutionary explanations of the human world. He starts his considerations with the fundamental differences that are made within everyday-lifeworld contexts by dealing with physical and living objects on the one hand and with autonomous actors on the other. According to differentiated grammar, the everyday lifeworld itself is articulated. This articulation is also relevant when we deal with evolutionary explanations. Those explanations necessarily refer to a situation. The initial description of the situation requires the application of a language, which is grammatically richer than the languages used as a basis for the explanation. Consequently, the transition from physical or chemical situations to situations of perception or moral judgment is grammatically (not empirically) excluded:

Entwicklungen in diesem Sinne können die Grammatik, in der die Anfangssituation und die Prinzipien ihrer Veränderung beschrieben werden, nicht verlassen. Von physikalischen und chemischen Zuständen und Ereignissen führt, weil kein
Following his argument, evolutionary discourse is understood as a narrative report of transformations from grammatically poorer to grammatically richer situations. As one important result of this approach, a shift from one language of description to another (with its respective unique grammar) cannot be conceptualized as a transition from one state to a following state of nature. The shift from one situation to another, understood as a real transition in the sense of an evolutionary transformation, can only be narrated:

- Die grammatisch neue Situation kann aus dem, was ihr vorausgeht, dann auch nicht erklärt oder verstanden werden. Zu ihren Ursprüngen kann man sich nur historisch erzählend im engeren Sinne verhalten, auch wenn sie, diese Ursprünge, der Ausgangspunkt neuer Entwicklungen werden, Entwicklungen, welche dann durchgängig eine grammatisch reichhaltigere Beschreibung verlangen. (Kambartel 1989 p 67)

Grammatical differentiation allows a powerful rejection of reductionist scientific approaches, not only within the narrow limits of evolutionary biology. Moreover, it provides a solution for our problem. If we have to state a multitude of situations that are “given” with a specific, adequate language of description and an adequate grammar, then a hiatus must be stated, that divides the human world from the world of animals, plants, physical and chemical objects. Consequently, connections between these different language-worlds can only be constructed by the composition of narrative explanations. However, we have to pay a high price for this liberation. If the connection between the different worlds can only take a narrative form, the relevance of scientific descriptions of humans, including evolutionary descriptions, becomes unclear. If no other connections exist than a mere coincidence between a specific property or faculty in one language-world and a specific set of events in another language-world, claims of validity of scientific descriptions and practices remain unsettled. The natural processes are separated from their respective description:

As a consequence of the plurality of language-worlds, the fundamental distinction between nature and culture remains intact. To avoid the resulting dualism, the relation between scientific description and the image of humans requires further consideration. The concept of human beings as persons having intentions, desires, or thoughts refers to a language of description that cannot be reduced to scientific descriptions:

► To think of a featherless biped as a person is to think of it as a being with which one is bound up in a network of rights and duties. From this point of view, the irreducibility of the personal is the irreducibility of the ‘ought’ to the ‘is’. But even more basic than this (…), is the fact that to think of a featherless biped as a person is to construe its behaviour in terms of actual or potential membership in an embracing group each member of which thinks of itself as a member of the group. (Sellars 1963 p 39)

According to Sellars, the combination of a scientific description of human beings as specimens and the description of human beings as persons transcends the limitations of a scientific explanation and finally results in a profound confusion of language levels. When we deal with humans as persons, the sprachspiel of intentions is assumed to be adequate:

► Now, the fundamental principles of a community, which define what is ‘correct’ or ‘incorrect’, ‘done’ or ‘not done’, are the most general common intentions of that community with respect to the behavior of the members of the group. It follows that to recognise a featherless biped or dolphin or Martian as a person requires that one think thoughts of the form, ‘We (one) shall do (or abstain from doing) actions of kind A in circumstances of kind C’. To think thoughts of this kind is not to classify or explain, but to rehearse an intention. (…) (Sellars 1963 p 40)

Sellars suggests a unity of anthropology by merging the concept of person with the scientific image of humans, thereby assuming that the resulting completion of the scientific image allows the revision of the dualism between man and nature:

► Thus the conceptual framework of person is not something that needs to be reconciled with the scientific image, but rather something to be joined to it. Thus, to complete the scientific image we need to enrich it not with more ways of saying what is the case, but with the language of community and individual intentions, so that by construing the actions we intend to do and the circumstances in which we intend to do them in scientific terms, we directly relate the world as conceived by scientific theory to our purposes, and make it our world
and no longer an alien appendage to the world in which we do our living. (Sellars 1963 p 40)

However, even if we accept the possibility of incorporating scientific descriptions of humans into the manifest image and adopting the resulting dialectics between both images, the central problem remains unsolved. It does not provide us with clues choosing among several possible descriptions, leading to different and sometimes even contradicting results. From this point of view, at the very least a prior theory of knowledge is required to reconstruct both natural and as social sciences. Sellars seems to be aware of the problem and concedes that the systematic relationship of nature and culture remains unclear. The dissolution of the resulting dualism therefore remains the work of the imagination:

- We can of course, as matters now stand, realise this direct incorporation of the scientific image into our way of life only in imagination. But to do so is, if only in imagination, to transcend the dualism of the manifest and scientific images of man-of-the-world. (Sellars 1963 p 40)

The dualism results from the fact that the grammatical objection itself does not permit us to determine the methodological status of scientific descriptions of human beings and, referring to those descriptions, the scientific explanation of human origins.

24.8 Conclusion I: the shortcoming of criteriology in paleoanthropology

We may achieve an adequate dissolution of this dualism if we take into account Cassirer’s definition of human beings by referring to their respective mediated activities. This most comprehensive definition provides us with the methodological starting point we are seeking. Referring to the respective human activities, we can give a biological description of aspects of these activities. Starting with body functions, such as the specific human mode of upright walking, the specific kind of human phonation, communication speech, etc., we are able to elaborate scientific descriptions that serve as the methodological starting points for the reconstruction of possible antecedent human constitutions. As we pointed out before, the resulting reconstructions allow us to reinterpret in terms of “possible living entities” (prehuman life-forms). According to Sellars, determination of the “images of man of the world,” two different language games, are the result of the scientific description of human beings:
1. The primary language game of our everyday lifeworld, which served as a methodological starting point and
2. The respective scientific language games, in terms, e.g., of biology, paleoanthropology, sociology, etc.

Taking into account that we need a third language game for life-scientific purposes, namely all the nonbiological, technical, physical, and chemical knowledge and know-how that we used in order to achieve our functional descriptions of living entities (in the as if mode), there are three fundamental asymmetries to consider:

1. Biological (genetic, physiological, morphological) descriptions of human beings (as if they were merely living entities), depends necessarily on nonbiological knowledge and know-how; in contrast, physics, technique, or chemistry are possible without any relevant reference to biology.
2. Paleoanthropology refers (besides other scientific sources such as geology, etc.) to biological knowledge (particularly in terms of functional descriptions) of recent life-forms. But on the other hand, this biological knowledge can be claimed to be valid without further reference to paleoanthropology.
3. Scientific descriptions of human beings refer necessarily to (self-) descriptions of human beings (images of the man in the world). Their validity and adequacy can be claimed only in the context of non- and prescientific practice and communication. Sciences are embedded into human culture—as an (specific and most relevant) aspect of this culture.

Considering this situation, we can give reasons for the source of the shortcomings of a criteriological foundation of paleoanthropology. When we deal with (nonhuman) living entities, we describe them, in the light of our self-knowledge, as integrating our technical and scientific knowledge and know-how. In the specific case we use as an explanatory framework here, we describe them “as if they were” functional units. The relation between the author (e.g., a scientist) and the objects (nonhuman living entities) of description changes in principle when author and object of description are both the same: now we describe ourselves as a certain kind of object, differing from other objects by some specific criteria. In our case, biological or paleoanthropological descriptions of human beings need to be considered as descriptions of human entities “as if they were living entities”—such as animals. To put this phenomenon of self-application with Husserl’s words:

▶ Zu der Problematik der Anomalität gehört auch das Problem der Tierheit und ihrer Stufenfolgen ‘höherer und niederer’ Tiere. In bezug auf das Tier ist der
Mensch, konstitutiv gesprochen, der Normalfall, wie ich selbst die Urnorm bin für alle Menschen; Tiere sind wesensmäßig konstituiert für mich also anomale ‘Abwandlungen’ meiner Menschlichkeit, mögen sich dann auch bei ihnen wieder Normalität und Abnormalität scheiden. (Husserl 1931 p 129)

If we overlook the inversion of the “as if” relation, a (possible and possibly valid) criteriological description gains the methodological status of a statement on the essence of human constitution. This indispensable asymmetry between a (poorer) scientific description and its (richer) methodical starting point has implications that are most relevant as the conditions of possibility of a paleoanthropological foundation of ethics.

24.9 Conclusion II: the criteriological failure of ethics

In criticizing criteriological approaches in ethics, we have reached a seemingly curious situation. On the one hand, the pitfalls of ethics can be recognized when we take into consideration the methodological problem with which paleoanthropological research is confronted. Here, we identified the necessity of starting with a semantically very rich description of human beings, their activities, and the very form of those activities, if only to gain a methodological starting point for a semantically poorer description of humans as a biological entity—named Homo sapiens. This leads finally to the difficulties we have identified of characterizing the transformation of a nonhuman into a human constitution as a transformation from nature to culture which remains nevertheless embedded into nature. Thus, the anticipation of the semantics of humanity is obscured by its translation into a biological language of characters. On the other hand, the normative deficiency of paleoanthropology becomes observable by explicating the semantics of a description of human beings and their activities in order to provide the basis for the fundament of ethical reasoning itself. Even by making a simple description of human beings as members of Homo sapiens, we claim at least its adequacy—and this claim cannot be replaced by an alternative description. So, even if we consider scientific theories to represent states of the world, at least the evolution of the adequacy of this representation cannot be given in a purely descriptive language again. This dilemma has its very origin in the fact of criteriological characterization of human beings, in both the fields of paleoanthropology and of ethical reasoning. And in both cases, we find shortcomings of criteriology, which leads us to the strong conclusion that on the one hand paleoanthropology and biology in general cannot provide any basis for ethical reasoning, and on the other hand those ethical approaches, which depend immanently on criteriological descriptions of humanity show serious methodological deficiencies.
References


Husserl E Cartesianische Meditationen (1931) In: Ders.: Gesammelte Schriften Bd. 8, Hamburg 1992


Pfordten vdD (1996) Ökologische Ethik, Rowohlt, München
Scheler M (1947) Die Stellung des Menschen im Kosmos. Nymphenburger Verlagshandlung, München