

## MOLECULAR EVOLUTION

V. A. RATNER

*Basing on the concrete examples of phylogenetic analysis of coded macromolecules (ribosomal RNAs, globins, influenza virus proteins, AIDS proteins, mt-DNA, etc.), the role of this new approach in modern development of Evolution Theory, Human Origin, and Humanity Evolution is shown.*

**На конкретных примерах филогенетического анализа кодируемых макромолекул (рибосомных РНК, глобинов, белков вируса гриппа, генов ВИЧ, мт-ДНК) показано, какую роль играет этот подход в современном развитии теории эволюции, происхождения человека и эволюции человечества.**

© Ратнер В.А., 1998

## МОЛЕКУЛЯРНАЯ ЭВОЛЮЦИЯ

В. А. РАТНЕР

Новосибирский государственный университет

Теория эволюции является одним из краеугольных камней биологии и естествознания в целом. В последние годы теория эволюции испытывала ускоренное изменение и развитие в связи с появлением в естественнонаучной практике нового мощного пласта экспериментальных данных и теоретических методов. В поле зрения теории эволюции попали кодирующие макромолекулы: ДНК, РНК и белки, составляющие костяк молекулярно-генетической организации клеток и организмов. Теперь уже широко известна роль ДНК и РНК как материальных носителей генов, м-РНК как переносчика генетической информации от генов к белкам и белков как исполнителей самых разнообразных биологических функций: ферментов — катализаторов процессов обмена, антител — иммунных защитников организма, структурных элементов, регуляторных и транспортных элементов и т.д.

Первые последовательности аминокислот в белках были расшифрованы (секвенированы) в 50-е годы, нуклеотидов в генах и РНК — в 60-е, но массовое секвенирование началось только в 1977–1978 годы, став к настоящему времени рутинной технологической операцией. Результаты секвенирования с тех пор накапливаются в компьютерных банках данных. К осени 1997 года крупнейшие международные банки данных содержали сведения примерно об одном миллиарде секвенированных нуклеотидов (~1, 6 млн генов и других фрагментов) и о более 100 тыс. белков.

Легко понять, что расшифровка нуклеотидных последовательностей генов (ДНК), РНК и аминокислотных последовательностей белков создает новую ситуацию, позволяющую взглянуть на процесс эволюции с молекулярного уровня организации жизни. Действительно, в классической генетике аллели генов принято обозначать однородными буквами или индексами: В и b, A1 и A2 и т.д. Если в процессе эволюции один аллель вытесняет другого (В ⇒ b и т.д.), то кажется, что происходят как бы утрата предыдущего аллеля (В) и распространение нового (b). В то же время в молекулярной генетике известно, что многие мутации состоят в замене лишь одного или нескольких нуклеотидов:

$$B \Rightarrow b$$

$$TTCGAGGCTAG \Rightarrow TTCGCGGCTAG$$

Поэтому замена аллеля может сводиться фактически к замене и фиксации единственного нуклеотида, а остальная, большая часть гена остается неизменной. Сходство последовательностей аллельных

генов (или белков) остается максимальным. Чем больше накапливается различий, тем меньше сходство последовательностей. Чем раньше дивергировали два гена, тем больше фиксированных различий они накопят.

Таким образом, сходство последовательностей макромолекул можно положить в основу молекулярно-филогенетических построений. Число различий мономеров (или близкая к нему величина) может быть мерой эволюционной дивергенции. Обилие секвенированных последовательностей в банках данных делает применимость этого подхода практически неограниченной.

В течение последних 30 лет были разработаны несколько десятков компьютерных методов и пакетов программ для построения филогенетических деревьев макромолекул и филогенетического анализа, при помощи которых были построены многие сотни филогенетических деревьев макромолекул. В результате их анализа получены многочисленные интереснейшие данные, причем некоторые из них имели фундаментальное значение. Многие макромолекулы эволюционировали гораздо медленнее, чем морфологические признаки живых форм, поэтому их филогенетический анализ позволяет заглянуть в очень ранние периоды эволюционного процесса (до 100–1000 млн лет тому назад).

На этом базируется развитие теории молекулярной эволюции. Здесь мы не будем описывать ее на

профессиональном уровне, этому посвящены другие издания, а продемонстрируем, так сказать, “научный продукт” этой науки – филогенетические деревья – и покажем, к каким любопытным выводам приводят их анализ и интерпретация. Некоторые приводимые результаты получены автором и его сотрудниками.

Кодирующие макромолекулы эволюционируют с разными скоростями. Наиболее консервативными обычно являются гены и белки некоторых очень глубоких и рано возникших генетических процессов, которые представлены у многих (или даже у всех) форм жизни. Таковы, например, гены рибосомных РНК (входящих в молекулярный механизм синтеза белка), некоторых гистонов – белков компактизации ДНК хромосом. Менее консервативны гены и белки систем, которые встречаются у определенных широких групп видов (например, глобины животных). Наконец, наиболее изменчивыми являются гены и белки РНК-содержащих вирусов, которые стремительно изменяются в борьбе с иммунной системой их хозяев (вирус гриппа, ВИЧ (HIV), онкогенные ретровирусы).

Примером глобального дерева, охватывающего все крупнейшие таксоны биологического мира, является дерево 16S–18S-рибосомных РНК, построенное американцем К. Вузом и сотрудниками (рис. 1). Наиболее неожиданным свойством этого дерева является существование в нем трех крупнейших

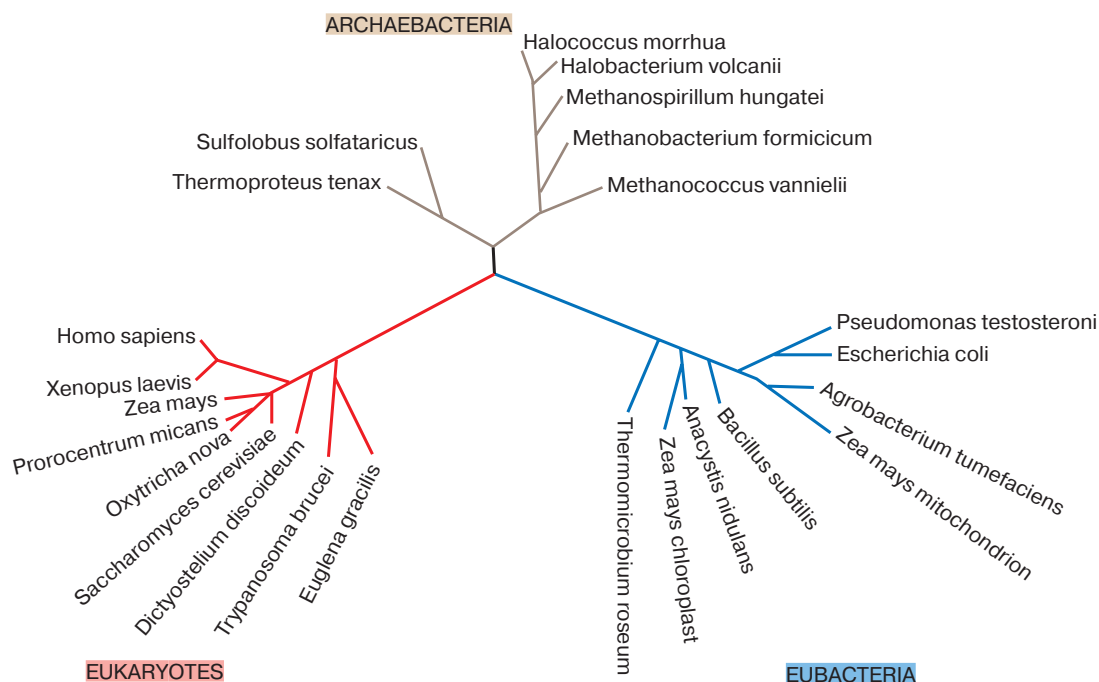


Рис. 1. Глобальное филогенетическое дерево надсемейства 16S–18S-рибосомальных РНК, построенное К. Вузом и сотрудниками

ветвей, объединяющих эубактерий (обычные бактерии), архебактерий (особая экзотическая группа бактерий) и эукариот (высшие ядерные организмы). До этой работы считалось, что бактерии составляют единое надцарство Prokariota, отделенное от второго надцарства — Eukariota существенной дистанцией. В дереве Вууза все три глобальные ветви примерно равно удалены друг от друга. Поэтому архебактерий следует считать самостоятельным надцарством. Таким образом, Вууз ввел три надцарства, или домена: Bacteria, Archaea, Eucarya. Некоторые исследователи предлагали также термин “империя”.

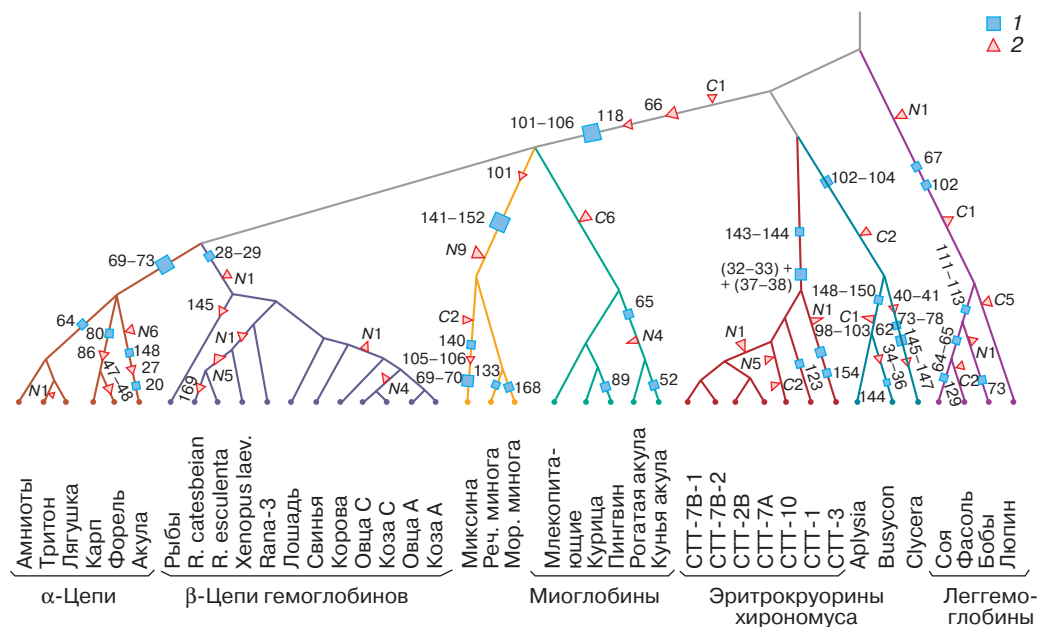
Этот важный результат сразу привлек внимание к изучению архебактерий, причем вскоре было показано, что по многим другим свойствам они действительно существенно удалены как от эукариот, так и от эубактерий. Многие архебактерии существуют в природе в весьма экзотических условиях: при высокой температуре вблизи подводных вулканов; в среде, насыщенной метаном, соединениями серы и т.д. Время разделения надцарств Bacteria и Archaea оценивается примерно в 3,5 млрд лет назад.

Дерево Вууза включает в себя фракции 16S-подобных р-РНК, кодируемых не только ядерными генами, но и генами клеточных органелл — митохондрий и хлоропластов. Легко заметить, что ядерная фракция р-РНК кукурузы (*Zea mays*) попадает в ветвь эукариот, как и ожидается, а фракции из митохондрий и хлоропластов этой же кукурузы (*Zea mays mitochondrion, chloroplast*) — в ветвь эубактерий. Этот факт считается самым веским аргументом

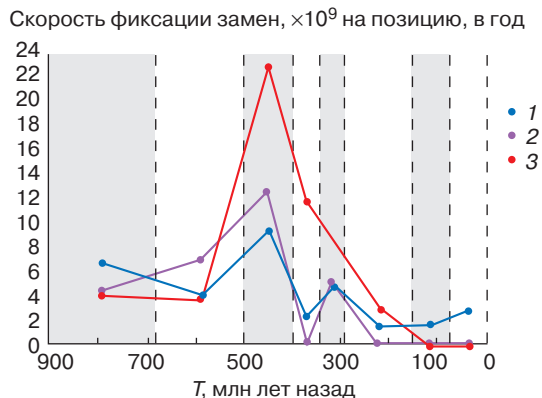
в пользу симбиотической гипотезы эволюционного возникновения эукариот, согласно которой митохондрии происходят от симбиотических предковых пурпурных бактерий, а хлоропласты — от цианобактерий (синезеленых водорослей), а вовсе не от ядерных структур эукариотических клеток.

Белки-глобины (гемоглобины крови, миоглобины тканей) и их гены распространены в живой природе не столь широко. Они найдены в основном у животных, некоторые удаленные их представители (леггемоглобины) найдены также у растений, а у прокариот они пока не обнаружены. Поэтому глобины образуют молекулярное надсемейство, охватывающее лишь часть таксономического дерева жизни. Деревья эволюции глобинов и их генов были построены многими исследователями. На рис. 2 приведено филогенетическое дерево глобинов, построенное А. Жарких в нашей лаборатории. В целом оно достаточно хорошо соответствует принятой таксономии животных.

Из независимых палеонтологических данных можно почерпнуть датировки эпох существования общих предков многих ныне живущих биологических форм. Так, в дереве глобинов датированы шесть точек ветвления. Поскольку методы реконструкции деревьев позволяют определить длины ребер дерева (числа фиксированных замен, мутаций) на различных маршрутах, то с учетом датировок это дает возможность подсчитать скорости эволюции на разных этапах и в разных ветвях дерева. Нами получены очень любопытные оценки (рис. 3).



**Рис. 2.** Филогенетическое дерево надсемейства глобинов, построенное А.А. Жарких для 40 удаленных последовательностей белков: 1 – фиксированные делеции, 2 – инсерции. Цифры указывают на их локализацию внутри белка

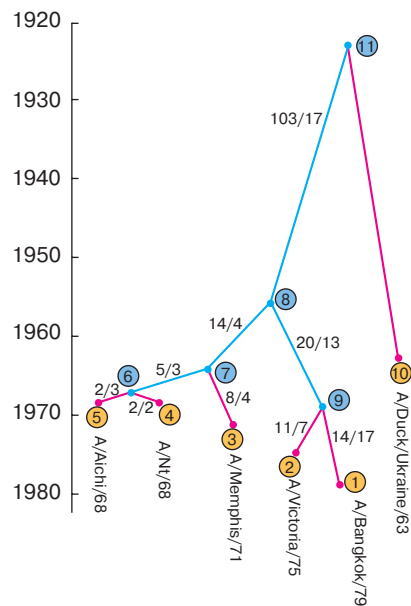


**Рис. 3.** Адаптивное ускорение фиксации замен аминокислот в гемоглобинах в эпоху выхода позвоночных животных на сушу (400–500 млн лет назад). Скорость эволюции: 1 – в среднем по белку; 2 – в центре контакта  $\alpha_1-\beta_2$ ; 3 – в регуляторном центре связывания ДФГ

Во-первых, оказалось, что средняя скорость эволюции по всему белку была непостоянна и имела максимум примерно 400–500 млн лет назад, в эпоху выхода позвоночных животных из Мирового океана на сушу (рис. 3, 1). Во-вторых, эта особенность выявилась еще ярче для тех участков глобинов, которые отвечают за образование их четвертичной (мультимерной, агрегатной) структуры. Дело в том, что миоглобины и гемоглобины некоторых примитивных животных до сих пор представляют собой протомеры (отдельные субъединицы белка). Круглоротые рыбы (например, минога) имеют димерные агрегаты гемоглобинов, а большинство других позвоночных – тетрамерные агрегаты, состоящие из двух  $\alpha$ -субъединиц и двух  $\beta$ -субъединиц ( $\alpha^2\beta^2$ ). Субъединицы соединяются при помощи так называемых центров контакта, которые обозначаются  $\alpha_1-\beta_1$ ,  $\alpha_1-\beta_2$  и т.д. Для современных гемоглобинов известны аминокислоты, входящие в центры контакта. Поэтому можно подсчитать скорости эволюции непосредственно для центров контакта. Максимум скорости четко выявляется для центра контакта  $\alpha_1-\beta_2$  (рис. 3, 2) и регуляторного центра связывания дифосфолипидерата (ДФЛ) (рис. 3, 3). Таким образом, при выходе позвоночных животных на сушу их гемоглобин приобрел тетрамерную структуру, когда разные субъединицы ( $\alpha$  и  $\beta$ ) связаны между собой центрами контакта  $\alpha_1-\beta_2$  и регуляторным центром ДФЛ. Прежде всего в эту эпоху (или немного раньше) ген общего предка гемоглобинов был дублирован, а его копии в ходе дивергенции дали начало двум родственным семействам:  $\alpha$  и  $\beta$ . Именно в ту эпоху мы видим максимальную скорость эволюции центров их контакта и регуляции, а затем в течение сотен миллионов лет скорость резко падала, часто до нуля, когда возникшие

центры были просто неизменны. Иначе говоря, мы как бы видим события эпохи формирования новых функциональных структур молекул гемоглобина, которые в дальнейшем сохраняются у всех наземных форм позвоночных. Следует учесть, что выход позвоночных на сушу и переход к дыханию свободным кислородом воздуха сопровождался резкой перестройкой всей системы дыхания, и в том числе структуры гемоглобинов. Ускорение эволюции в эту эпоху означает, что указанные приобретения были высоко адаптивными, то есть обеспечивали существенное преимущество их обладателям.

Примером дерева наиболее быстрой молекулярной эволюции является дерево генов, кодирующих гемагглютинины Н3 вируса гриппа (рис. 4), построенное Л.В. Омелянчуком и сотрудниками. Это белок вирусного капсида, некоторые выпуклые участки которого (антигенные детерминанты) узнаются специфическими антителами хозяина (человека). В результате иммунное сопротивление хозяина препятствует размножению вируса. Вирус гриппа имеет РНК-геном, в котором мутации происходят с наибольшей известной частотой  $\sim 10^{-3}-10^{-4}$  на позицию, за репликацию. В XX веке изучены несколько локальных эпидемий и пандемий гриппа. Одной из



**Рис. 4.** Филогенетическое дерево генов гемагглютинаина Н3 вируса гриппа типа А, построенное Л.В. Омелянчуком и др. «Висячие» вершины и ведущие к ним ребра отвечают эпидемическим штаммам (1, 2, 3, 4, 5, 10). Индексы штаммов указывают на место и год выделения. Внутренние вершины и ребра дерева отвечают неэпидемическим стадиям циркуляции вируса («резервуар»). Вблизи ребер указаны числа синонимических/несинонимических замен нуклеотидов

первых изученных была пандемия испанки 1918–1919 годов. В дальнейшем были зафиксированы пандемии гонконгского гриппа и др. В большинстве случаев образцы эпидемических штаммов вируса гриппа были собраны и сохранены в коллекциях. После 1978 года РНК этих эпидемических штаммов были секвенированы, а позже теоретиками построены филогенетические деревья.

Анализ дерева на рис. 4 показал, что все эпидемические варианты оказались на нижних тупиковых ветвях дерева (так называемые висячие вершины). Это значит, что эпидемические штаммы, как правило, не бывают прямыми потомками других эпидемических штаммов. Иначе говоря, неэпидемический вирус, видимо, циркулирует в каком-то “резервуаре” (локальной популяции человека, где к вирусу есть устойчивый иммунитет, или в в популяции животных, которые могут быть промежуточными носителями вируса). Эпидемии возникают тогда, когда появляется новый вариант вируса, против которого нет готовых антител, то есть готового иммунитета хозяина. Этот вариант как бы выплескивается в человеческую популяцию и быстро в ней распространяется, пока не будет выработан адекватный иммунный ответ. Тогда эпидемия затухает. Следующая эпидемия возникает путем независимого выплескивания нового варианта вируса гриппа, не связанного, как правило, с предыдущей эпидемией.

Поскольку все эпидемии были датированы (см. рис. 4), оказалось возможным определить скорости эволюции. В эпидемических ветвях они оказались в 3–5 раз выше, чем в неэпидемических, причем особенно большое ускорение (в 36 раз) было выявлено именно в позициях, кодирующих антигенные детерминанты белка. Иммунное сопротивление хозяина размножению вируса является главным селективным фактором, действующим на вирус. Поэтому выплеск эпидемического штамма, не встречающего иммунного сопротивления хозяина, сопровождается ускорением эволюции. Это размножение высоко адаптивно для вируса и неадаптивно для хозяина-человека. Напротив, после выработки иммунного ответа спокойное размножение вируса становится для него и хозяина нейтральным — в ожидании новых адаптивных мутаций и рекомбинаций. В целом можно говорить о коэволюции вируса и иммунного ответа системы хозяина. Вирус стремится как бы мутационно выскользнуть из-под готового иммунного ответа хозяина, а хозяин стремится догнать новые варианты вируса путем выработки нового специфичного иммунного ответа.

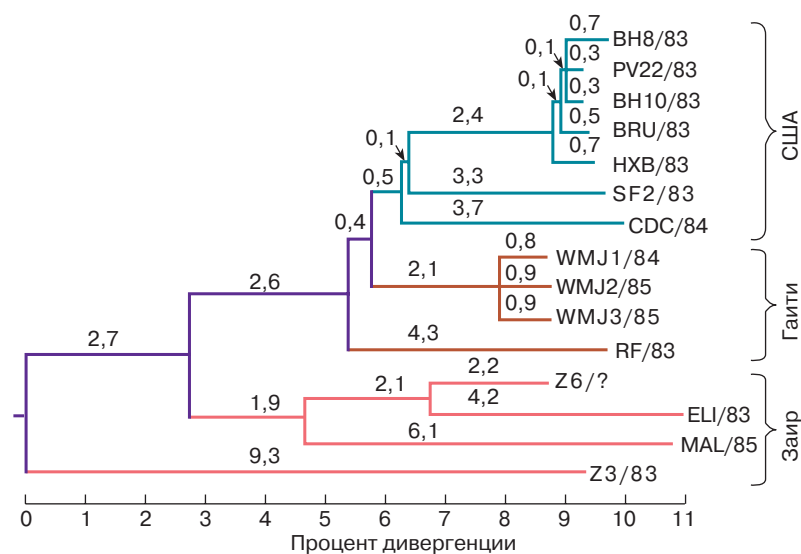
Еще один пример — эволюция вируса ВИЧ (вирус иммунодефицита человека, HIV в английской транскрипции), вызывающего заболевание СПИД (синдром приобретенного иммунодефицита, в английской транскрипции AIDS). Вирус ВИЧ передается половым путем, а также нередко при переливании крови, нестерильных инъекциях и хирургических

вмешательствах, при родах от матери к ребенку и в других подобных ситуациях. Он был обнаружен в начале 80-х годов в крови больных, погибавших от различных тяжелых заболеваний. Эти заболевания имели одно общее свойство — иммунная система больного была неспособна противостоять как этому вирусу, так и сопровождавшим возбудителям других заболеваний, от которых, собственно, больные и погибали. Были обследованы образцы из банков переливания крови, которые созданы во многих странах для медицинских целей. Оказалось, что многие образцы содержали вирус ВИЧ.

Различные штаммы вируса ВИЧ, выделенные в разных районах Африки, Карибского бассейна и США, были секвенированы, для них теми же методами, что и выше, построены филогенетические деревья (одно из них показано на рис. 5). Это дерево указывает, что вирус ВИЧ существовал в Центральной Африке (Заир) до 1960 года, был занесен на Гаити до середины 70-х годов и в США до 1978 года. Иначе говоря, его истоки лежат в Африке, где в некоторых странах вирусоносители составляют до половины населения, хотя заболевание СПИД проявляется далеко не у всех. Видимо, размножение вируса все же сдерживается их иммунной системой. Вирус ВИЧ сходен с некоторыми вирусами обезьян, поэтому многие ученые предполагают, что он возник в результате изменчивости этих вирусов и был занесен в популяцию человека извне относительно недавно.

Интересно, что изменчивость вируса ВИЧ (мутационная и рекомбинационная) столь велика, что он успевает измениться в сторону усиления своей агрессивности непосредственно в ходе развития болезни СПИД у некоторых больных в течение 1,5–2 лет и менее. Оценки изменчивости показывают, что она превышает верхний допустимый предел, в результате чего иммунная система больного не успевает справиться с выработкой новых вариантов антител и фактически распадается, открывая путь процессу заболевания СПИД и других болезней. Именно в этом, по-видимому, состоит источник агрессивности заболевания СПИД.

Есть биологические проблемы, которые в большой степени и всегда интересуют человека и человечество. Среди них особое место занимает проблема происхождения и расово-этнической эволюции самого человечества. В последние годы в этой области прогресс связан в основном с филогенетическим анализом макромолекул. Американский ученый М. Гудмен и сотрудники построили филогенетические деревья для некоторых генов и белков от высших приматов и человека. На рис. 6 приведено одно из них. Из него следует, что порядок последовательного ответвления видов от эволюционного ствола человека таков: макак резус — орангутанг — горилла — шимпанзе — человек. Иначе говоря, наиболее близким к человеку является шимпанзе.



**Рис. 5.** Филогенетическое дерево штаммов вируса ВИЧ (HIV), вызывающего заболевание СПИД, построенное В.-Т. Ли и сотрудниками. Числа вблизи ребер указывают процент дивергенции последовательностей. Справа указаны регионы, где изолированы штаммы

Следует отметить, что до этих работ в таксономии высших приматов выделяли два семейства: Hominidae, включавшее только вид *Homo sapiens*, и Pongidae, включавшее гориллу, шимпанзе и орангутанга. Молекулярно-филогенетический анализ существенно изменил эту классификацию. Гудмен и его сотрудники предложили новую таксономию высших приматов:

#### Семейство Hominidae

##### Подсемейство Homininae

Род *Gorilla* (1 вид – горилла),

Род *Homo* (1 вид – человек),

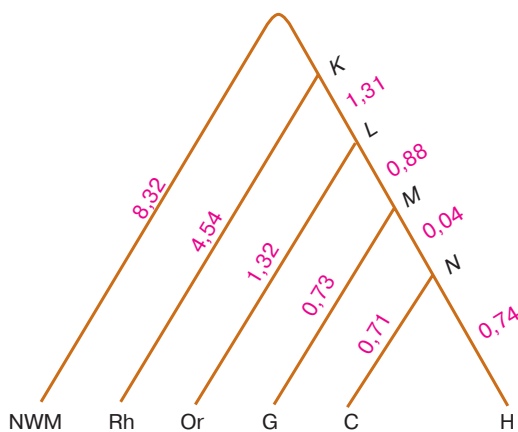
Род *Pan* (2 вида: обычный шимпанзе и карликовый шимпанзе),

##### Подсемейство Ponginae

Род *Pongo* (1 вид – орангутанг)

Близость человека и шимпанзе дала основание Дж. Даймонду назвать человека третьим шимпанзе. По его оценкам, геномы человека и шимпанзе различаются примерно на  $3,2 \cdot 10^7$  одиночных замен нуклеотидов из  $3 \cdot 10^9$ , то есть на каждый сотый нуклеотид, но остальные 99 из 100 нуклеотидов у них одинаковы. Время их молекулярной дивергенции оценивается в 5–10 млн лет назад. Эти факты являются очень веским аргументом в пользу естественного происхождения человека в процессе филогенетического развития жизни на Земле в противовес различным религиозным мифам о разовом творении 5–10 тыс. лет тому назад.

Таким образом, молекулярно-филогенетический анализ позволяет приблизиться к решению очень важных вопросов эволюции. Иногда это единственная возможность заглянуть далеко назад. При этом почти всегда мы можем оперировать только кодирующими последовательностями от ныне живущих видов. “Ископаемых” последовательностей почти нет. Правда, в последнее время удается выделить и секвенировать фрагменты ДНК из замороженных ископаемых животных (мамонт), мумифицированных тел человека. Особых открытий эти



**Рис. 6.** Филогенетическое дерево высших приматов, построенное М. Гудменом и сотрудниками для некодирующих последовательностей из района вокруг гена ф $\eta$ -глобина. Обозначения: Н – человек, С – шимпанзе, G – горилла, Or – орангутанг, Rh – макак резус, NWM – обезьяны Нового Света. Цифры вблизи ребер означают числа замен на 100 позиций нуклеотидов

данные пока не содержат. Исключение представляет уникальная возможность выделения ДНК из мелких насекомых (мухи, термиты, пчелы, долгоносик), законсервированных в природной смоле деревьев (янтарь). Известно, что эту идею использовал американский кинорежиссер С. Спилберг в фильме “Парк юрского периода”: в янтаре якобы найден комар, в кишечнике которого содержатся остатки клеток крови динозавра, укушенного этим комаром; из них выделяется ДНК динозавра. Затем динозавр полностью воспроизводится генно-инженерными методами на радость и страх посетителям парка. Разумеется, это фантазия. Но анализ ДНК законсервированных насекомых вполне реален. Так, для термита, пчелы и долгоносика из янтара были клонированы и секвенированы гены 18S-рибосомальной РНК. Они включены в филогенетические деревья этих молекул. Возраст найденных в янтаре термитов и пчел – 25–40 млн лет, долгоносика – 120–135 млн лет. Пока это рекордные случаи.

В заключение можно отметить, что молекулярный филогенетический анализ становится сейчас одним из весомых методов таксономии живых форм. Таксономия высших приматов сейчас больше базируется на молекулярных данных, чем на каких-либо других. В систематике наивысших таксонов, систематике высших растений решающую роль играют деревья рибосомных РНК. Таксономию бактерий вообще невозможно построить без молекулярных данных. Можно надеяться, что в дальнейшем роль этого подхода только возрастет.

## ЛИТЕРАТУРА

1. Эволюция. М.: Мир, 1981.
2. Айала Ф. Введение в популяционную и эволюционную генетику. М.: Мир, 1984.
3. Айала Ф., Кэйгер Дж. Современная генетика. М.: Мир, 1988. Т. 3.
4. Кимура М. Молекулярная эволюция: Теория нейтральности. М.: Мир, 1985.
5. Ратнер В.А., Жарких А.А., Колчанов Н.А. и др. Проблемы теории молекулярной эволюции. Новосибирск: Наука, 1985.
6. Ратнер В.А. Краткий очерк теории молекулярной эволюции. Новосибирск: НГУ, 1992.
7. Ratner V.A., Zharkikh A.A., Kolchanov N.A. et al. Molecular Evolution. В. etc.: Springer-Verlag, 1996.

\* \* \*

Вадим Александрович Ратнер, доктор биологических наук, профессор Новосибирского государственного университета, зав. теоретическим отделом Института цитологии и генетики СО РАН, действительный член РАЕН. Область научных интересов: математическая генетика, теория молекулярной эволюции, теория молекулярно-генетических систем. Автор более 300 научных публикаций, в том числе 14 монографий.