ГЕНЕТИКА, 2005, том 41, № 12, с. 1–10

ГЕНЕТИКА
МИКРООРГАНИЗМОВ

УДК 575.116.4

ГЕННЫЕ ПОРЯДКИ В ГЕНОМАХ АЛЬФАПРОТЕОБАКТЕРИЙ: СХОДСТВО И ЭВОЛЮЦИЯ

© 2005 г. И. А.Захаров[1], А. В. Марков[2]

[1] Институт общей генетики им. Н.И.Вавилова Российской академии наук, Москва 119991;
e-mail: zakharov@@vigg.ru

[2] Палеонтологический Институт Российской академии наук, Москва 117647;
e-mail: markov–a@inbox.ru

Поступила в редакцию 25.07.2005 г.

Генные порядки в геномах девяти альфа-протеобактерий и трех гамма-протеобактерий сравнивали при помощи количественных показателей *S* (относительное число общих пар соседних генов) и L (средняя разность межгенных расстояний). Использовалась выборка из 200 гомологичных генов, присутствующих у всех 11 штаммов. Во всех проанализированных геномах обнаружено 20 консервативных “неразрывных” участков, включающих в общей сложности 63 гена из 200. Скорость изменений порядка генов в эволюции сильно варьирует в разных эволюционных линиях. Наибольший темп изменений характерен для внутриклеточного паразита *Wolbachia* (*Rickettsiales*) (40–60 геномных перестроек за 100 млн. лет). При помощи компьютерного моделирования показано, что наблюдаемое в исследуемой выборке соотношение величин *S* и L свидетельствует о несколько сниженной вероятности крупных геномных перестроек по сравнению с мелкими.

Эволюция геномов совершается на двух уровнях – на уровне нуклеотидных последовательностей и на уровне порядков генов в хромосомах. Если замены нуклеотидов существенным образом не понижают жизнеспособность, они сохраняются; в противном случае они отметаются естественным отбором. Накопление в геномах в процессе эволюции замен нуклеотидов следует правилу молекулярных часов и именно поэтому сравнение последовательностей ДНК успешно используется при реконструкции филогении организмов. Изменение генных порядков происходит в результате инверсий, транслокаций и транспозиций. Нет убедительных свидетельств того, что порядок генов в хромосомах контролируется естественным отбором. Если бы темп перестройки хромосом в ходе эволюции был постоянным, филогению организмов можно было бы реконструировать, сравнивая их генные порядки, т.е. их генетические карты.

Впервые один из авторов настоящей работы [1] предложил количественно измерять сходство генетических карт. На основе разработанных мер сходства было проведено сравнение геномов млекопитающих, показана возможность реконструкции филогении геномов [2.3] и промоделирована эволюция геномов, состоящих из нескольких хромосом [4.5]. Использованные в этих работах меры сходства предназначались для сравнения геномов, в которых неизвестен порядок генов в хромосомах, а известно только их расположение в той или иной хромосоме. Была предложена также и мера сходства собственно генных порядков, которая давала возможность сравнивать геномы, представленные не группами синтении (порядок генов в которых неизвестен), а группами сцепления [6].

Для подобных сравнений другие меры были разработаны Санковым и соавт. [87, 8], которые сравнили генные порядки в митохондриях разных организмов. В дальнейшем этим автором и его коллегами было проанализировано изменение генных порядков и в хромосомах млекопитающих [9].

Количественные методы при сравнении генных порядков в геномах бактерий были впервые применены Суияма и Борком [10]. Недавно была опубликована работа Белда и соавт. [11], в которой проанализирована филогения генных порядков у гамма-протеобактерий.

В настоящей работе, взяв за основу более ранние исследования [1–6], мы попытались проанализировать эволюцию хромосом некоторых представителей альфа-протеобактерий, для которых известна полная нуклеотидная последовательность геномов, что дало нам возможность сравнивать локализацию гомологичных (ортологичных) генов. Особенно нас интересовали представители порядка Rickettsiales, в который входят внутриклеточные бактерии – симбионты членистоногих и патогены млекопитающих.

Таблица 1. Бактерии, геном которых анализировался в настоящей работе

Вид	Штамм	Размер генома, пн	Систематическое положение	Обозначение генома	Номер в базе данных*
Wolbachia pipientis	endosymbiont of *Drozophila melanogaster*	1267782	Rickettsiales, Wolbachieae	wMel	NC_002978
Wolbachia sp.	endosymbiont of *Brugia malayi* str. TRS	1080084	То же	wBm	NC_006833
Rickettsia conorii	str. Malish 7	1268755	Rickettsiales, Rickettsieae	rc	NC_003103
Rickettsia prowazekii	str. Madrid E	1111523	То же	rp	NC_000963
Rickettsia typhi	str. Wilmington	1111496	»	rt	NC_006142
Anaplasma marginale	str. St. Maries	1197687	Rickettsiales, Anaplasmataceae	an	NC_004842
Ehrlichia ruminantum	str. Welgevonden	1512977	То же	eh	NC_006832
Rhodopseudomonas palustris	CGA009	5459213	Rhizobiales, Bradyrhizobiaceae	rhp	NC_005296
Escherihia coli	CFT073	5231428	Enterobacteriales, Enterobacteriaceae	ec	NC_004431
Salmonella enterica subsp. *enterica*	serovar. Typhi Ty2	4791961	То же	sa	NC_004631
Shigella flexneri	2a str. 301	4607203	»	sh	NC_004337

* http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?val=

МАТЕРИАЛЫ И МЕТОДЫ

Геномы

Сравнивали порядки генов в геномах восьми альфа-протеобактерий и взятых для сравнения трех гамма-протеобактерий, перечисленных в табл. 1. Для анализа использовали 200 гомологичных генов, каждый из которых присутствует в каждом из 11 геномов в единственном числе (гены, дуплицированные хотя бы в одном из геномов, были исключены из рассмотрения). Список включенных в анализ генов представлен на сайте http://macroevolution.narod.ru / supplem01.htm

Сравнение генных порядков

Для сравнения генных порядков использовались перечисленные ниже количественные показатели.

1. Относительное число общих пар соседних генов в двух геномах: $S = m/n$, где m – число одинаковых пар соседних генов в двух сравниваемых геномах (соседними гены считались в том случае, если между ними в хромосоме нет ни одного гена из рассматриваемой выборки, включающей 200 генов); n – общее число пар соседних генов в одном геноме (в данном случае $n = 200$). Данная мера эквивалентна мере, предложенной в работе [6]. В тех случаях, когда вместо меры сходства нужно было использовать меру различия, вместо S применялась величина $1 - S$.

2. Средняя разность расстояний между генами (L). Показатель вычисляли следующим образом. Для каждого гена вычислялось расстояние до всех остальных генов в данном геноме; всего, таким образом, для генома определялось $(200^2 - 200)/2 = 19900$ межгенных расстояний. Каждое межгенное расстояние вычислялось как модуль разности условных номеров позиций двух генов в геноме (условный номер позиции – порядковый номер гена в последовательности из 200 рассматриваемых в данной работе генов); с учетом того, что хромосома кольцевая, величина межгенного расстояния x, превышающая 200/2 = 100, заменялась на величину $100 - x$. Затем вычислялось абсолютное значение разности межгенных расстояний для каждой из 19900 пар гомологичных генов в двух сравниваемых геномах. Итоговая величина L вычислялась как среднее из этих абсолютных значений.

3. Мера, основанная на принципе парсимонии. Наличие или отсутствие тех или иных пар генов-соседей у разных видов можно использовать как независимые признаки для построения кладограмм, основанных на принципе парсимонии (строится филогенетическое дерево, предполагающее наименьшее возможное число “событий” – переходов признака из одного состояния в другое, в нашем случае – появлений и исчезновений конкретных пар генов-соседей в геноме).

Таблица 2. Относительное число общих пар соседних генов (мера межвидового сходства, *S*)

Геномы	wBm	rc	rp	rt	an	eh	rhp	ec	sa	sh
wMel	**0.520**	0.245	0.245	0.245	0.265	0.265	0.255	0.250	0.250	0.245
wBm		0.245	0.250	0.250	0.265	0.265	0.275	0.245	0.245	0.240
rc			**0.975**	**0.965**	0.260	0.260	0.285	0.265	0.270	0.260
rp				**0.990**	0.260	0.260	0.285	0.265	0.270	0.260
rt					0.260	0.260	0.285	0.265	0.270	0.260
an						**0.780**	0.265	0.250	0.250	0.255
eh							0.265	0.255	0.255	0.255
rhp								**0.350**	**0.355**	**0.360**
ec									**0.840**	**0.885**
sa										**0.835**

Примечание. Видовую принадлежность геномов см. в табл. 1. Жирным выделены значения, превышающие фоновый уровень.

Моделирование эволюции генных порядков

Для анализа полученных результатов была разработана компьютерная модель, имитирующая пошаговое изменение порядка генов во времени.

В модели задается исходный произвольный порядок из 200 генов. Затем на каждом шаге модели происходит одно “событие”: транслокация или инверсия. Каждое событие моделируется следующим образом:

1) выбирается первая точка разрыва;

2) выбирается вторая точка разрыва;

3) кольцевой геном разделяется на две части;

4) большая из двух частей замыкается в кольцо и в ней выбирается третья точка разрыва;

5) меньшая часть встраивается в третью точку разрыва в прямом или обратном порядке.

Модель позволяет задавать различные опции, или входные параметры эксперимента:

1. Вероятность разрыва может быть одинакова во всех 200 возможных точках, а может варьировать. Имеется возможность запретить разрывы в некоторых точках (создать консервативные “неразрывные” участки в геноме), а также задать пониженную вероятность разрыва в определенных точках (создать консервативные участки с низкой вероятностью разрыва).

2. Расстояние между первой и второй точками разрыва может быть случайным (т.е. определяться как случайное целое число в диапазоне от 1 до 100), а может быть ограничено. Во втором случае пользователь задает отдельно вероятность того, что расстояние от точки 1 до точки 2 будет равно 1 (вырежется только один ген), 2–3 (вырежется 2–3 гена), 4–6, 7 или более.

3. Расстояние от вырезанного участка (между точками 1 и 2) до третьей точки разрыва может быть случайным, а может ограничиваться. Во втором случае отдельно задается вероятность того, что это расстояние будет нулевым (в этом случае вырезанный кусок встроится на то же место, из которого был вырезан, с определенной вероятностью инверсии), равным 1, 2–3, 4–6, 7 и более.

4. Задается вероятность инверсии, т.е. того, что вырезанный участок будет встроен в хромосому в обратном порядке.

На каждом шаге модели измененный геном сравнивается с исходным и вычисляются величины *S* и L. Некоторые результаты моделирования при разных входных параметрах показаны на графиках (см. ниже).

РЕЗУЛЬТАТЫ

Сравнение генных порядков

Результаты сравнения генных порядков 11 штаммов при помощи показателей *S* и *L* представлены в табл. 2, 3.

На основе табл. 2 и 3 при помощи программы Fitch из пакета Phylip 3.6 (Fitch-Margoliash method) были построены деревья (дендрограммы), отражающие соотношение величин *S* и *L* у рассматриваемых 11 видов бактерий (рис. 1,*а, б*). При помощи программы Pars из пакета Phylip 3.6 (Wagner parsimony method) была построена кладограмма (рис. 1,*в*), основанная на таблице бинарных признаков, каждый из которых отражает присутствие (1) или отсутствие (0) конкретной пары генов-соседей в данном геноме. Исходные данные для построения кладограммы представлены на сайте: http://macroevolution.narod.ru / supplem01.htm#pars

Динамика показателей S и L в модельных экспериментах

Динамика величины S в целом более устойчива и менее зависит от входных параметров и от

Таблица 3. Средняя разность расстояний между генами (мера межвидовых различий, L)

Геномы	wBm	rc	rp	rt	an	eh	rhp	ec	sa	sh
wMel	32.04	31.57	31.59	31.50	31.29	31.12	32.30	32.00	31.58	32.01
wBm		31.80	31.84	31.94	31.83	31.77	32.04	32.08	32.50	32.12
rc			**00.79**	**02.46**	30.52	31.34	31.70	32.89	32.57	32.80
rp				**01.74**	30.47	31.34	31.70	32.86	32.54	32.78
rt					30.28	31.34	31.76	32.89	32.61	32.84
an						**16.03**	32.30	32.30	31.91	32.21
eh							31.94	32.52	31.74	32.44
rhp								32.02	30.95	31.76
ec									**17.99**	**03.46**
sa										**18.19**

Примечание. Видовую принадлежность геномов см. в табл. 1. Жирным выделены значения, превышающие фоновый уровень.

фактора случайности, чем L. Величина S в ходе модельной "эволюции" убывает с замедлением и стабилизируется на некотором уровне. Входные параметры сравнительно мало влияют на скорость убывания и хаотичность динамики S (степень непредсказуемости, разброса результатов при одних и тех же параметрах) (рис. 2).

Наиболее существенно влияет на динамику S внесение в модельный геном "неразрывных" участков. Если их не задавать (допустить равновероятные разрывы во всех точках), то S в течение 200–400 шагов стабилизируется на уровне, характерном для "полного хаотического перемешивания" (порядка 0.02–0.04) (рис. 2,*а*). Если же задать "неразрывные участки" точно такие же, как в наших данных (см. ниже – одна группа из 20 сцепленных соседних генов, одна из пяти, одна из четырех и 17 групп из двух), то S стабилизируется примерно на том уровне, который является "фоновым" в нашей выборке (порядка 0.22–0.26). Если дополнительно задать еще 10 точек с пониженной в 2–3 раза вероятностью разрыва, уровень стабилизации S совпадает с наблюдаемым фоновым уровнем еще точнее (0.24–0.27) (рис. 2,*б, в*).

Поведение величины L в модельных экспериментах более разнообразно и хаотично. В целом величина L возрастает с замедлением и стабилизируется на одном и том же, почти не зависящем от параметров устойчивом уровне (порядка 31–33), что соответствует "полному перемешиванию". В отличие от величины "уровня стабилизации" скорость возрастания величины L и степень хаотичности ее динамики очень сильно зависят от входных параметров, а именно от соотношения вероятностей мелких и крупных геномных перестроек (вероятности "вырезания" малых и больших участков генома и перемещения их на малые или большие расстояния).

В зависимости от этих параметров для величины L возможны три альтернативных режима изменений во времени:

1. Если длина вырезаемого участка и расстояние, на которое он перемещается, задаются случайно (т.е. вероятность крупных перестановок не ниже, чем мелких), величина L возрастает очень быстро и достигает максимума всего за 10–40 шагов модели. Хаотичность при этом невелика (т.е. результаты разных последовательных "прогонов" модели при одних и тех же параметрах сравнительно мало отличаются друг от друга) (рис. 2,*а*).

2. Если вероятность вырезания небольшого участка генома выше, чем крупного, и (или) вероятность его перенесения на небольшое расстояние выше, чем на далекое, величина L растет медленнее, стабилизируется на максимальном уровне позже, а степень хаотичности в динамике L оказывается очень высокой: разные последовательные "прогоны" модели при одних и тех же входных параметрах могут весьма сильно отличаться друг от друга (рис. 2,*б*). Это объясняется тем, что крупные геномные перестройки приводят к резкому скачкообразному росту величины L. Поэтому хаотичность динамики L невелика в двух диаметрально противоположных ситуациях: когда крупные перестройки часты и регулярны (рис. 2,*а*) и когда они не происходят вовсе (рис 2,*в*). Если же они происходят редко и нерегулярно, динамика L становится малопредсказуемой (рис. 2,*б*). При дальнейшем снижении вероятности крупных геномных перестроек рост L продолжает замедляться; хаотичность вначале остается высокой, затем постепенно снижается.

3. Если задать крайне низкую или нулевую вероятность крупных геномных перестроек (например, полностью запретить вырезание крупных участков и(или) перенос их на большое расстояние), рост L очень сильно замедляется; в резуль-

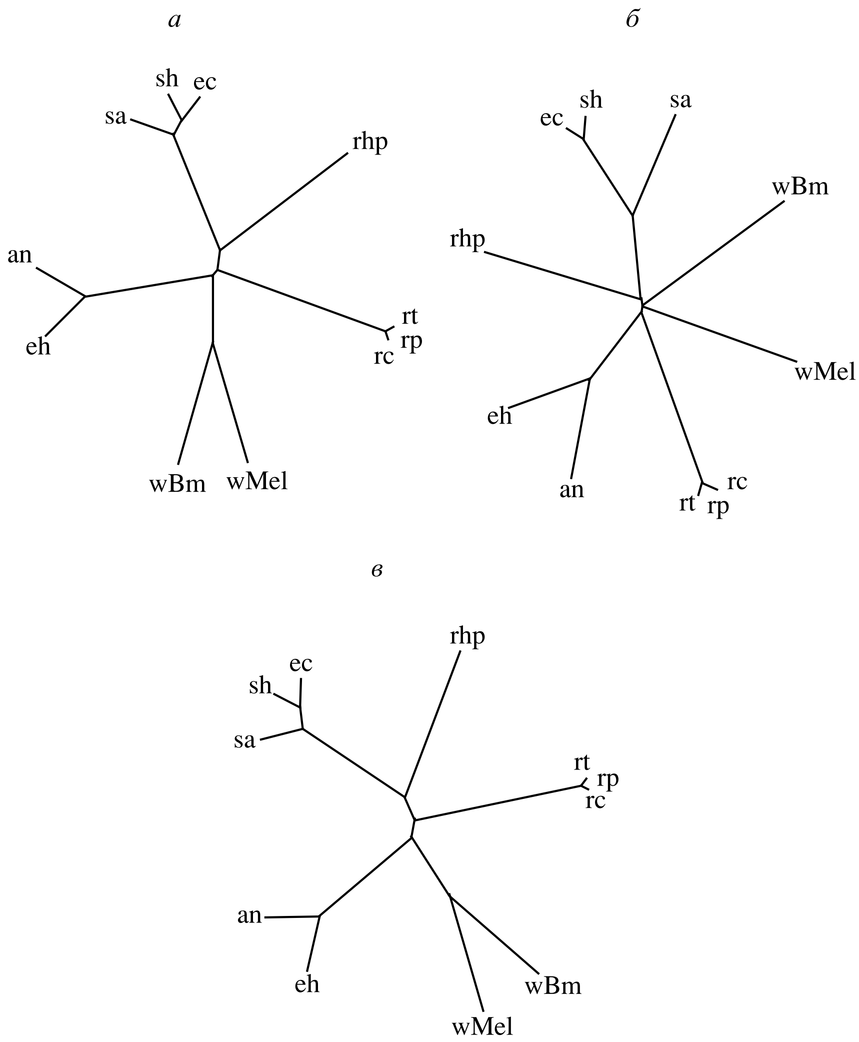

Рис. 1. Деревья, отражающие сходство генных порядков у 11 исследованных штаммов. *а* – дерево, построенное на основе показателя 1 – S (табл. 2); *б* – дерево, построенное на основе показателя L (табл. 3); *в* – кладограмма, построенная на основе принципа парсимонии (исходные данные представлены на сайте http://macroevolution.narod.ru/supplem01.htm#pars.

тате этот показатель может достигать стабильного уровня много позже, чем *S*. Хаотичность динамики *L* при этом резко снижается (рис. 2,*в*).

ОБСУЖДЕНИЕ

Сравнение геномов при помощи показателя S

Как указывалось, использованная нами при анализе сходства бактериальных геномов мера *S* эквивалентна мере, предложенной в работе [6], в которой с ее помощью определялось сходство геномов, представленных не одной кольцевой хромосомой, а несколькими группами сцепления. В работе Бланчетте и соавт. [8] та же мера (без ссылки, однако, на публикацию [6]) использована при сравнении кольцевых митохондриальных геномов; авторы показали, что, несмотря на свою простоту, она дает лучшие результаты, чем другие, более математически сложные меры.

В настоящем исследовании показатель *S* позволил выделить в пределах рассматриваемой выборки из 11 штаммов бактерий пять четко обособленных кластеров (рис. 1): 1) rc + rp + rt, 2) wMel + wBm, 3) an + eh, 4) rhp, 5) ec + sa + sh. Такое разделение согласуется с принятой таксономической схемой; кластеры соответствуют таксонам: Rickettsieae, Wolbachieae, Anaplasmataceae, Bradyrhizobiaceae, Enterobacteriaceae.

Степень сходства между кластерами, определяемая по величине *S*, примерно одинакова для всех пар кластеров и составляет от 0.245 до 0.360 (тогда как минимальное сходство между штаммами внутри кластера 0.520). Наибольшее сходство наблюдается между кластерами 4 и 5 (0.350–0.360), во всех остальных парах кластеров сходст-

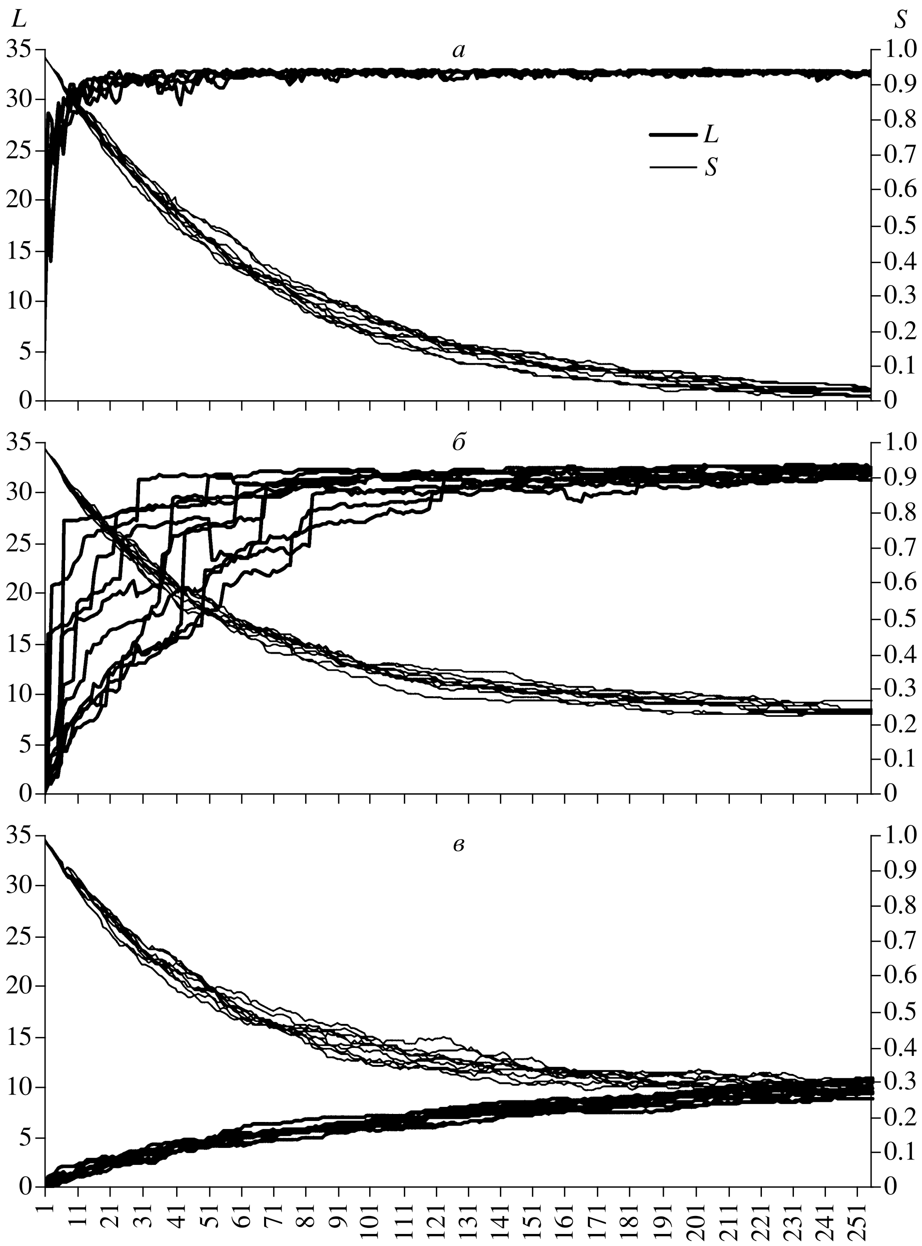

Рис. 2. Изменение параметров S и L в модельных экспериментах. На каждой фигуре показаны результаты 10 реализаций (прогонов) модели. По горизонтальной оси – шаги модели ("события"). Входные параметры модели: вероятность инверсии вырезанного куска 0.5; *а* – "неразрывные" участки генома отсутствуют, длина вырезаемого участка случайная (любая), расстояние переноса случайное (любое); *б, в* – заданы "неразрывные" участки генома (такие же, как в исследуемой выборке), а также 10 позиций с пониженной в 2.5 раза вероятностью разрыва: *б* – вероятность того, что длина вырезаемого участка составит один ген, равна 0.7, 2–3 – 0.2, 4–6 – 0.06, 7 или более – 0.04, расстояние переноса случайное, т.е. имеются "неразрывные" участки, мелкие перестройки генома более вероятны, чем крупные; *в* – вероятность того, что длина вырезаемого участка составит один ген, равна 0.3, 2–3 – 0.3, 4–6 – 0.4, 7 или более – 0; вероятность переноса на нулевое расстояние – 0, на 1-ю позицию – 0.2, 2–3-ю – 0.4, 4–6-ю – 0.4, 7-ю или более – 0, т.е. крупные перестройки генома запрещены.

во находится в пределах 0.245–0.285. Эти результаты не соответствуют принятой классификации (сходство между кластерами является почти постоянной величиной и не пропорционально степени родства, отраженной в таксономической удаленности). В противном случае следовало бы ожидать, что представители Rickettsiales (кластеры 1, 2 и 3) будут более сходны друг с другом, чем с 4-м и 5-м, и т.д. Причина несоответствия заключается, очевидно, в том, что закономерности изменения генных порядков в ходе эволюции существенно отличаются от закономерностей изменения нуклеотидных последовательностей внутри генов.

Почти одинаковая степень сходства генных порядков, определяемая по величине S, между всеми пятью кластерами (притом, что таксономические и филогенетические расстояния между ними очень сильно различаются) позволяет предположить, что в ходе дивергенции эволюционных линий сравнительно быстро достигается некий предельный уровень "перемешанности" генов, по достижении которого дальнейшее уменьшение S тормозится или вовсе останавливается.

Можно предположить три альтернативных объяснения наблюдаемой картины:

1. Достижение полной хаотической перемешанности. В этом случае величины S, характеризующие степень сходства удаленных клад, должны быть близки к тем, которые получаются в модельном эксперименте при сравнении двух абсолютно случайных генных порядков.

2. Возможно, в геноме существует определенное количество устойчивых блоков сцепленных генов (в частности, оперонов), для которых вероятность фиксации разрывов очень мала или равна нулю, т.е. сцепленное состояние этих генов поддерживается стабилизирующим отбором. В этом случае величины S, характеризующие степень сходства удаленных клад, должны соответствовать полному хаотическому перемешиванию всего генома, за исключением этих консервативных "групп сцепления". Эти величины будут существенно больше, чем получаемые при сравнении двух абсолютно случайных генных порядков. При этом они могут быть приблизительно одинаковы в самых разных парах сравниваемых клад, если набор консервативных "групп сцепления" в этих кладах один и тот же.

3. Возможно и иное объяснение наблюдаемого одинакового уровня сходства между кластерами. Можно предположить, что консервативные блоки сцепленных генов отсутствуют или малочисленны, максимально возможная перемешанность еще не достигнута, а наблюдаемое сходство величин S во всех парах рассматриваемых кластеров возникло случайно и объясняется тем, что перемешивание генома происходило с разными скоростями в разных кладах.

Наблюдаемая картина полностью соответствует второму из трех приведенных выше предположений. При полном хаотическом перемешивании 200 генов в кольцевой хромосоме величина S, отражающая сходство двух геномов, с вероятностью $p < 0.01$ не будет превышать 0.030; с вероятностью $p < 0.05$ она не будет превышать 0.025 (цифры получены в результате компьютерного моделирования по методу Монте-Карло). Это соответствует 5–6 одинаковым парам соседних генов (из 200) в двух сравниваемых геномах. Такое число одинаковых пар соседних генов обеспечивается простой случайностью. Мы же имеем на порядок больший "фоновый" уровень сходства между удаленными кладами (0.245–0.360, что соответствует 49–72 одинаковым парам из 200). Это позволяет отбросить гипотезу 1.

Чтобы проверить гипотезу 3, необходимо посмотреть, какими общими парами соседних генов обеспечивается "фоновый" уровень сходства между кластерами. Если эти пары генов будут одни и те же во всех парах сравниваемых геномов, то это будет подтверждать гипотезу 2, если разные – гипотезу 3.

В действительности оказалось, что в нашей выборке присутствуют 20 консервативных групп сцепления, общих для всех без исключения рассматриваемых видов, относящихся как к альфа-, так и к гамма-протеобактериям. Одна из этих групп включает 20 генов, другая – пять, третья – четыре и 17 групп включают по два сцепленных гена. Ниже приводится список этих "групп сцепления":

1) preprotein translocase, SecY subunit + ribosomal proteins L15 + S5 + L18 + L6 + S8 + S14 + L5 + L24 + + L14 + S17 + L16 + S3 + L22 + S19 + L2 + L23 + L4 + + L3 + S10;

2) ribosomal protein L11 + ribosomal protein L1 + + ribosomal protein L10 + ribosomal protein L7 / L12 + + DNA-directed RNA polymerase, beta / beta' subunits;

3) NADH dehydrogenase I, J subunit + NADH dehydrogenase I, K subunit + NADH dehydrogenase I, L subunit + NADH dehydrogenase I, M subunit;

4) ribosomal protein L19 + tRNA (guanine-N1)-methyltransferase;

5) translation elongation factor Ts + ribosomal protein S2;

6) ribosomal protein L9 + ribosomal protein S18;

7) ribosomal protein S12 + ribosomal protein S7;

8) ribosome recycling factor + uridylate kinase;

9) ribosomal protein L13 + ribosomal protein S9;

10) heat shock protein HslV + heat shock protein HslVU, ATPase subunit HslU;

11) NADH dehydrogenase I, B subunit + NADH dehydrogenase I, A subunit;

12) Holliday junction DNA helicase RuvA + Holliday junction DNA helicase RuvB;

13) succinyl-CoA synthetase, alpha subunit + succinyl-CoA synthase, beta subunit;

14) ribosomal protein S15 + polyribonucleotide nucleotidyltransferase;

15) chaperonin, 60 kDa + chaperonin, 10 kDa;

16) ribosomal protein L17 + DNA-directed RNA polymerase, alpha subunit;

17) ribosomal protein S11 + ribosomal protein S13;

18) NADH dehydrogenase I, H + NADH dehydrogenase I, G subunit;

19) ATP synthase F1, alpha subunit + ATP synthase F1, delta subunit;

20) glycyl-tRNA synthetase, beta subunit + glycyl-tRNA synthetase, alpha subunit.

В общей сложности эти консервативные “неразрывные” участки генома включают 43 пары соседних генов, присутствующие у всех 11 штаммов. Таким образом, только одни эти консервативные участки обеспечивают уровень $S = 0.215$. С учетом того, что при полном хаотическом перемешивании еще около пяти общих пар сцепленных генов обеспечиваются простой случайностью (что соответствует увеличению S еще на величину порядка 0.02–0.03), мы видим, что наблюдаемый “фоновый” уровень величины S, по-видимому, полностью объясняется существованием указанных “неразрывных” участков генома.

“Неразрывные” участки в большинстве своем состоят из генов, близких по функции. Это, в первую очередь, гены белков, связанных с механизмом трансляции (в том числе рибосомные белки); во многих случаях “неразрывные группы сцепления” включают гены субъединиц одного и того же белкового комплекса (например, NADH-дегидрогеназы). Это могут быть опероны или просто гены функционально связанных белков, которые должны экспрессироваться совместно, в постоянном количественном соотношении, и синхронно попадать (доставляться) в одни и те же части клетки. “Неразрывность” группы генов-соседей может поддерживаться отбором благодаря большей эффективности координированной экспрессии и (или) транспортировки мРНК при сближенном положении этих генов в хромосоме. Очевидно, что эти блоки унаследованы от общего предка альфа- и гамма-протеобактерий и сохранились в эволюции.

По-видимому, выявленные в исследуемой выборке пять “кластеров” близких видов уже удалились друг от друга слишком сильно, чтобы величину их сходства или родства можно было оценивать при помощи величины S. Они уже достигли в своем расхождении стабильного максимального уровня перемешанности геномов, абсолютная величина которого (S порядка 0.25) определяется наличием определенного числа “неразрывных” участков, и дальнейшие “перетасовки” их геномов уже не могут снизить этот уровень сходства.

Сравнение геномов при помощи показателя L

Показатель L позволяет выделить в пределах рассматриваемой выборки из 11 штаммов бактерий не пять (как в случае с показателем S), а шесть кластеров (табл. 3, рис. 1). Один из S-кластеров, включающий два штамма *Wolbachia*, распался на два L-кластера, по одному штамму в каждом. Так же, как и в первом случае (при сравнении на основе показателя S), между всеми кластерами наблюдается примерно одинаковый (“фоновый”) уровень различий (L от 30.28 до 32.89). Максимальный уровень различий внутри кластера 18.19 (между *Salmonella* и *Shigella*).

При полном хаотическом перемешивании геномов в 5000 экспериментов Монте-Карло величина L не выходит за пределы 32.42–33.43. С вероятностью $p < 0.01$ величина L оказывается более 32.72, с вероятностью $p < 0.05$ – более 32.89. Таким образом, наблюдаемый “фоновый” уровень различий лишь ненамного меньше уровня “случайного шума”.

По-видимому, это объясняется тем, что показатель L отражает “глобальную” степень перемешанности, и сохранение отдельных небольших “неразрывных” участков генома лишь ненамного снижает его величину.

В целом можно заметить положительную корреляцию показателей $1 - S$ и L: в тех парах штаммов, в которых один из этих показателей низкий, второй обычно тоже имеет небольшую величину, и наоборот. Однако в ходе эволюции, по-видимому, показатель L быстрее достигает уровня, близкого к предельному (т.е. характерному для полной хаотической перемешанности, с учетом присутствия “неразрывных” участков), и потому обладает меньшей разрешающей способностью.

У двух штаммов *Wolbachia*, по сравнению с членами всех остальных кластеров, оказался наибольший уровень различия генных порядков (по обоим показателям: $1 - S$ и L). Величина различия генных порядков у двух штаммов вольбахии превышает не только межвидовые различия у ближайших родственников вольбахии – риккетсий, но и межродовые различия между *Anaplasma* и *Ehrlichia* и между *Escherichia, Shigella* и *Salmonella*. Это свидетельствует о том, что перемешивание генома в ходе эволюции вольбахии происходило значительно быстрее, чем в других группах, в том числе близкородственных. Это хорошо согласуется с тем фактом, что в геноме вольбахии обнаружено аномально большое число мобильных генетических элементов [12].

Кладистический анализ

На основе данных о наличии и отсутствии конкретных пар соседних генов в геномах была получена единственная наиболее парсимоничная кладограмма, показанная на рис. 1,*в*. И положение, и относительная длина ветвей оказались практически идентичными у этой кладограммы и у дерева, построенного по матрице межвидовых “расстояний”, оцененных при помощи показателя $1 - S$ (табл. 2, рис. 1,*а*). При этом надо учесть, что объем информации, используемой при построении кладограммы, значительно больше, чем при построении дерева по матрице различий (расстояний) (рис. 1,*а*). В первом случае учитывался каждый признак (наличие или отсутствие данной пары генов-соседей) по отдельности, и находилось такое филогенетическое дерево, в котором число эволюционных событий (приобретений или утрат пар генов-соседей) было бы минимальным. Использовалась матрица из 9779 нулей и единиц (http://macroevolution.narod.ru/supplem01.htm#pars). Во втором случае использовалась обобщенная мера различия (число пар генов-соседей, присутствующих в одном геноме и отсутствующих в другом); анализируемая матрица включала лишь 55 значимых чисел (табл. 2).

Сравнение модельных результатов с фактическими данными

Сравнение результатов компьютерного моделирования с фактическими данными позволяет сделать вывод о характере эволюции генных порядков в рассматриваемой группе. Наблюдаемая в реальности (в исследуемой выборке) картина лучше всего соответствует тем модельным экспериментам, в которых задавалось наличие "неразрывных" участков генома (т.е. сниженная или нулевая вероятность разрыва для определенных позиций) и умеренное, но не слишком резкое ограничение вероятности крупных геномных перестроек (эта ситуация отражена на рис. 2,*б*). Действительно, только при таких входных параметрах можно получить те соотношения S и L, которые наблюдаются в исследуемой выборке, а именно: L на уровне "шума" (31–32) при S около 0.5 или менее; $L = 16–18$ при $S = 0.78 – 0.84$; $L = 3.5$ при $S = 0.89$; $L = 1 – 2.5$ при $S = 0.96–0.99$.

Из этого можно заключить, что в эволюции рассматриваемых групп бактерий крупные геномные перестройки происходили с существенно меньшей частотой, чем мелкие, но все же вероятность крупных перестроек была достаточно высокой.

Итак, проведенное нами сравнение геномов нескольких протеобактерий показало, что темп хромосомной эволюции (перестройки генных порядков) совершенно различен в разных группах. Особенно он высок в роде *Wolbachia* и более низок у других представителей порядка *Rickettsiales*. Обращает на себя внимание тот факт, что геномы всех этих внутриклеточных симбионтов или паразитов утратили сколько-нибудь заметное сходство с геномом свободно живущего *Rhodopseudomonas*. При этом геном последнего вида, относимого к альфа-протеобактериям, сохранил некоторое сходство с геномами гамма-протеобактерий (табл. 2).

Анализ эволюции генных порядков в группе гаммапротеобактерий был недавно выполнен Белда и соавт. [11], которые применили другие меры измерения сходства генных порядков, чем те, которые использовали мы. На основании проведенного анализа авторам удалость построить дендрограммы сходства геномов 30 видов, по топологии подобные тем, которые получаются при сравнении нуклеотидных последовательностей. В анализ были включены три вида симбионтов насекомых – *Buchnera aphidicola, Wigglesworthia glossinidia, Blochmannia floridanus*. Оказалось, что скорость перестройки генома у разных эндосимбионтов различна: у *Buchnera* и *Wigglesworthia* она выше, а у *Blochmannia* такая же, как и у свободно живущих бактерий. Авторы делают вывод, что эндосимбиоз не всегда связан с повышенным темпом перестроек генома. Наши результаты не противоречат такому заключению. Исключительно высокая лабильность генома *Wolbachia* представляется пока уникальным явлением.

Известно, что две линии *Wolbachia*, к которым принадлежат штаммы *wMel* и *wBm* (соответственно, симбионты насекомых и нематод), в эволюции разошлись около 100 млн. лет назад [13]. Результаты проведенного нами компьютерного моделирования показывают, что за это время их геномы стали различными в результате возникновения и фиксации не менее чем 40–50 перестроек. Столь высокий темп изменения порядка генов в ходе эволюции этих внутриклеточных альфа-протеобактерий резко контрастирует с пониженным темпом перестроек генома митохондрий у большинства Metazoa (митохондрии являются потомками симбиотических альфа-протеобактерий). Например, порядок митохондриальных генов у всех хордовых, дивергенция которых началась свыше 500 млн. лет назад, практически одинаков (различия не превышают одной-двух инверсий или транспозиций) [8]. Это различие в темпах изменения порядка генов дополняет список ключевых различий митохондрий и вольбахий, приведенный нами ранее [14].

Работа финансово поддержана грантом Подпрограммы № 2 Программы № 25 Президиума РАН "Происхождение и эволюция биосферы".

СПИСОК ЛИТЕРАТУРЫ

1. *Захаров И.А., Валеев А.К.* Количественный анализ эволюции геномов млекопитающих посредством сравнения генетических карт // Докл. АН СССР. 1988. Т. 301. С. 1213–1218.
2. *Захаров И.А., Никифоров В.С., Степанюк Е.В.* Гомология и эволюция генных порядков: простой метод проверки гипотез о характере этой эволюции // Генетика. 1996. Т. 32. № 1. С. 128–132. (*Zakharov I.A., Nikiforov V.S., Stepanyuk E.V.* Homology and evolution of gene orders: A simple method for test-

ing a hypothesis on the pattern of this evolution // Rus. J. Genetics. 1996. V. 32. № 1. P. 112–116.)

3. *Zakharov I.A.* Measurements of similarity of synteny groups and an analysis of genome rearrangements in the evolution of mammals // Bioinformatics and Genome Research: Proc. 3d Intern. Conf. / Ed. Lim H.A., Cantor C.R. Singapore: World Sci. Publ. Co, 1995. P. 107–113.
4. *Захаров И.А., Никифоров В.С., Степанюк Е.В.* Гомология и эволюция генных порядков: комбинаторная мера сходства групп синтении и моделирование эволюционного процесса // Генетика. 1992. Т. 28. № 1. С. 77–81.
5. *Захаров И.А., Никифоров В.С., Степанюк Е.В.* Гомология и эволюция генных порядков: моделирование и реконструкция эволюционного процесса // Генетика. 1997. Т. 33. № 1. С. 31–39. (*Zakharov I.A., Nikoforov V.S., Stepanyuk E.V.* Homology and evolution of gene orders: Simulation and reconstruction of the evolutionzry process // Rus. J. Genetics. 1997. V. 33. № 1. P. 24–30.)
6. *Захаров И.А., Никифоров В.С., Степанюк Е.В.* Измерение сходства порядков гомологичных генов // Генетика. 1991. Т. 27. № 2. С. 367–369.
7. *Sankoff D., Leduc G., Antoine N. et al.* Gene order comparisons for phylogenetic inference: Evolution of mitochondrial genome // Proc. Natl Acad. Sci. USA. 1992. V. 89. P. 6575–6579.
8. *Blanchette M., Kunisawa T., Sankoff D.* Gene order breakpoint evidence in animal mitochondrial phylogeny // J. Mol. Evol. 1999. V. 49. P. 193–203.
9. *Ehrlich J., Sankoff D., Nadeau J.H.* Synteny conservation and chromosome rearrangements during mammalian evolution // Genetics. 1997. V. 147. № 1. P. 289–296.
10. *Suyama M., Bork P.* Evolution of prokaryotic gene order: Genome rearrangements in closely related species // Trends Genet. 2001. V. 21. P. 10–13.
11. *Belda E., Moya A., Silva F.J.* Genome rearrangement distances and gene order phylogeny in gamma-Proteobacteria // Mol. Biol. Evol. 2005. V. 22. P. 1456–1467.
12. *Wu M., Sun L.V., Vamathevan J. et al.* Phylogenomics of the reproductive parasite Wolbachia pipientis wMel: A streamlined genome overrun by mobile genetic elements // PLoS Biol. 2004. V. 2. № 3. P. 327–341.
13. *Bandi C., Anderson T.J.C., Genchi C., Blaxter M.L.* Phylogeny of Wolbachia-like bacteria in filarial nematodes // Proc. R. Soc. London. Ser. B. 1998. V. 265. P. 2407–2413.
14. *Захаров И.А., Марков А.В.* Внутриклеточный паразит *Wolbachia* и происхождение эвкариот // Успехи соврем. биологии. 2005. Т. 125. С. 323–336.