

МОЛЕКУЛЯРНАЯ ОСНОВА ЗАКОНА РЯДОВ ГОМОЛОГИЧЕСКОЙ ИЗМЕНЧИВОСТИ Н.И. ВАВИЛОВА

И.Б. Рогозин^{1,2}, В.И. Глазко³, Е.В. Кунин¹

¹Национальный информационный центр по биотехнологиям NLM, Национальный институт здоровья, Бетезда, США, e-mail: rogozin@mail.nih.gov; ²Институт цитологии и генетики СО РАН, Новосибирск, Россия; ³РГАУ-МСХА им. К.А. Тимирязева, Москва, Россия

Редкие геномные мутации (rare genomic changes – RGCs) часто являются высокоинформативными признаками, надежно характеризующими таксоны на молекулярном уровне. Такие признаки играют все большую роль в реконструкции «дерева жизни», так как считается, что они меньше подвержены влиянию основного артефакта филогенетического анализа – гомоплазии. Нами был предложен новый вариант RGC, обозначенный как RGC_CAM (консервативные множественные аминокислотные замены – Conserved Amino acids-Multiple substitutions), учитывающий такие аминокислотные замены X > Y, которые: а) локализованы в гомологичной позиции множественного выравнивания аминокислотных последовательностей, кодируемых ортологичными генами, б) в исследуемом выравнивании наблюдаются два типа аминокислот (X и Y), аминокислота Y присутствует в одном или двух родственных таксонах, аминокислота X характеризует более широкий спектр таксонов, и в) переход из X в Y требует двух или трех нуклеотидных замен. В данной работе проведен анализ уровня гомоплазии RGC_CAM, возникающей в результате параллельных аминокислотных замен в 462 генах из 19 эукариотических видов. Обнаружено, что такие параллельные изменения наблюдаются много чаще в ветвях филогенетического дерева, которые более близки по времени дивергенции к предковому виду по сравнению с более удаленными ветвями. Это соответствует закону Н.И. Вавилова о гомологических рядах в наследственной изменчивости и согласуется с коварионной моделью молекулярной эволюции.

Введение

Закон гомологических рядов в наследственной изменчивости Николая Ивановича Вавилова в значительной степени лег в основу формирования новой области исследований – эволюционной генетики. Суть и уникальность его заключаются в выявлении связи между параллелизмом рядов изменчивости и близостью происхождения разных видов и родов или, другими словами, генетической обусловленности фенотипической изменчивости в различных таксономических группах (Vavilov, 1922).

Явление параллелизма привлекало внимание многих естествоиспытателей, в том числе и Ч. Дарвина, который писал: «Под этим термином (аналогичной или параллельной изменчивостью) я разумею, что одинаковые признаки время от времени (occasionally) проявляются у некоторых разновидностей или рас, ведущих на-

чало от одного и того же вида, и, реже, в потомстве отдаленных видов... Случаи аналогичной изменчивости в отношении их происхождения могут быть подразделены на две категории: во-первых, на случаи, зависящие от неизвестных причин, действующих на одинаковые конституции, а отсюда и варьирующие одинаково, и, во-вторых, на такие случаи, которые обязаны выявлению признаков более или менее отдаленных предков» (Цит. по: Вавилов, 1935).

Параллелизм изменчивости признаков объединяет целый ряд феноменов, начиная с присутствия у разных таксонов истинно общих (полученных от общего предка) генетических признаков, а также независимые друг от друга события: собственно параллельные изменения, обратные замены (реверсии) и конвергентную изменчивость (Юрченко, Захаров, 2007). Эта сложность приводила к критике закона Н.И. Вавилова (например, Филипченко, 1924). Очевид-

но, что только сравнительная оценка вклада в наблюдаемый параллелизм общности генетических основ и фиксации случайных независимых изменений в связи с их относительной приспособленностью может позволить выявить механизмы возникновения параллелизма. Одним из наглядных примеров сложности такой задачи могут быть современные работы по поискам и картированию главных генов количественных признаков, в которых достаточно часто обнаруживается, что одинаковое проявление одного и того же признака может быть обусловлено полиморфизмом разного набора главных («критических») для этого признака генов (например, Sinha *et al.*, 2006).

Сложность механизмов, лежащих в основе параллелизма изменчивости в близких таксонах, может приводить к неоднозначности в филогенетических построениях. Несмотря на развитие методов молекулярной генетики, до сих пор не удалось количественно описать параллельные замены на молекулярном уровне. В настоящее время сравнение секвенированных последовательностей геномов разных видов широко используется для филогенетического анализа и становится основной стратегией реконструкции эволюционной истории живых организмов (Wolf *et al.*, 2002; Snel *et al.*, 2005). Редкие геномные отличия (rare genomic changes – RGC), уникальные для отдельных таксонов, являются перспективными филогенетическими маркерами (Rokas, Holland, 2000; Nei, Kumar, 2001; Delsuc *et al.*, 2005; Boore, 2006). RGC соответствуют необратимым филогенетическим маркерам классической кладистики (синапоморфы или маркеры Хеннига) (Hennig, 1950; Rokas, Holland, 2000; Boore, 2006). RGC выбираются таким образом, чтобы отличия между ними были результатом единичных (или нескольких) редких мутационных событий $X > Y$, таких, что обратная мутация $Y > X$ была крайне маловероятной или даже невозможной. Нами недавно был предложен новый вариант RGC, обозначенный как RGC_CAM (множественные консервативные аминокислотные замены – Conserved Amino acids-Multiple substitutions), который основан на сравнительном анализе аминокислотных последовательностей (Rogozin *et al.*, 2007a). RGC_CAM подход предполагает выявление таких позиций в ами-

нокислотных последовательностях, которые консервативны в исследуемом таксоне (например, эукариоты) за исключением одного или нескольких видов (например, позвоночные), отличающихся присутствием в гомологичных позициях других аминокислот, объединяющих их в группу (рис. 1). RGC_CAM анализ предполагает, что характеристика, общая для основных эукариотических линий (растения, животные, грибы и *Apicomplexa*), является предковой, а отличающаяся от нее и выявленная у отдельной группы видов имеет более позднее происхождение (Rogozin *et al.*, 2007a). Анализ RGC_CAM выполняется с использованием статистических методов, позволяющих сравнить набор филогенетических гипотез и выбрать наиболее надежную в отношении оценки кладистической иерархии видов (Rogozin *et al.*, 2007a).

Hs	MSLIC	<u>S</u>	N	EVPEHPCVSPVS ...
Mm	MSLIC	<u>S</u>	N	EVPEHPCVSPVS ...
Dm	MALV	<u>C</u>	N	EVPETPVVSPHS ...
Ag	MSLV	<u>C</u>	N	EVPEHPCISPKS ...
Ce	MSFV	<u>C</u>	G	ELTEDPVVSQVS ...
Cb	MSFV	<u>C</u>	G	EPTEDPVVSPVS ...
At	M--	<u>N</u>	G	EVPEEPVVSCKS ...
Sc	M--	<u>L</u>	G	KVPRRPVLSPKS ...
Sp	M--	<u>F</u>	G	ETPKPEVISRVS ...
Pf	MSII	<u>C</u>	G	QTPEEPVIS-KT ...

Рис. 1. Пример RGC_CAM (подчеркнуты).

Сокращения: *Homo sapiens* (Hs), *Caenorhabditis elegans* (Ce), *Drosophila melanogaster* (Dm), *Saccharomyces cerevisiae* (Sc), *Schizosaccharomyces pombe* (Sp), *Arabidopsis thaliana* (At), *Anopheles gambiae* (Ag), *Plasmodium falciparum* (Pf), *Caenorhabditis briggsae* (Cb), and *Mus musculus* (Mm), *Brugia malayi* (Bm), *Aedes aegypti* (Aa), *Ciona intestinalis* (Ci), *Apis mellifera* (Am), *Cryptococcus neoformans* (Cn), *Dictyostelium discoideum* (Dd), *Nematostella vectensis* (Nv), *Strongylocentrotus purpuratus* (St) и *Trichoplax adhaerens* (Ta).

Существенной проблемой для применения методов RGC в эволюционных исследованиях являются параллельные изменения, т. е. фиксация одних и тех же аминокислотных замен в разных линиях, не обусловленных наличием общего предшественника и возникающих независимо друг от друга вследствие случайных причин (Telford, Budd, 2003). Метод RGC_CAM способствует уменьшению количества ошибок

в оценках гомоплазии, поскольку учитывает только такие аминокислотные замены, которые возникают в результате двух или трех нуклеотидных замен. Множественные нуклеотидные замены редки, так что вероятность внесения ошибок в оценки гомоплазии существенно ниже, чем при включении в анализ аминокислотных замен, возникающих при однонуклеотидных отличиях (Averof *et al.*, 2000; Silva, Kondrashov, 2002). Однако это не гарантирует полного отсутствия таких ошибок при использовании RGC_CAM метода (Rogozin *et al.*, 2007a). В этой связи нами разработана схема прямой оценки возможного количества параллельных нуклеотидных замен, приводящих к появлению RGC_CAM. Оказалось, что такие параллельные замены могут возникать много чаще во внутренних ветвях отдельных таксономических единиц, более близких к общему предку по сравнению с отличиями между таксономическими единицами, расхождение которых является относительно более поздним событием. Полученные данные на молекулярном уровне соответствуют известному закону Н.И. Вавилова о гомологических рядах в наследственной изменчивости, а также коварионной модели молекулярной эволюции.

Материалы и методы

462 выравнивания консервативных белков, использованных в анализе, были сформированы на основании базы данных ортологичных генов euKaryotic Orthologous Groups (KOGs) (Tatusov *et al.*, 2003) и включали ортологичные гены 10 эукариотических видов с полностью секвенированными геномами: *Homo sapiens* (Hs), *Caenorhabditis elegans* (Ce), *Caenorhabditis briggsae* (Cb), *Drosophila melanogaster* (Dm), *Saccharomyces cerevisiae* (Sc), *Schizosaccharomyces pombe* (Sp), *Arabidopsis thaliana* (At), *Anopheles gambiae* (Ag), *Plasmodium falciparum* (Pf), *Mus musculus* (Mm) (Rogozin *et al.*, 2007a). Эти последовательности доступны на сайте ftp://ftp.ncbi.nlm.nih.gov/pub/koonin/RGC_CAM/. К этим последовательностям с использованием программы COGNITOR (Tatusov *et al.*, 2003) были добавлены ортологичные последовательности из других эукариотических геномов: *Brugia malayi*

(Bm), *Aedes aegypti* (Aa), *Ciona intestinalis* (Ci), *Apis mellifera* (Am), *Cryptococcus neoformans* (Cn), *Dictyostelium discoideum* (Dd), *Nematostella vectensis* (Nv), *Strongylocentrotus purpuratus* (St), *Trichoplax adhaerens* (Ta). Для того чтобы уменьшить проблемы, связанные с ошибками выравнивания последовательностей, в анализ включали только те, выравнивание которых было однозначным и не вызывало сомнений: все участки, содержащие делеции или инсерции даже в одной последовательности, удалялись из анализа вместе с 5 прилегающими позициями (Rogozin *et al.*, 2007a, b). Метод RGC_CAM подробно описан в статье И.Б. Рогозина и соавторов (Rogozin *et al.*, 2007a).

Результаты

Филогенетические взаимоотношения между животными, рассмотренными в работе, представлены на рис. 2. Длина ветвей оценена в единицах RGC_CAM метода. Обнаружено, что три нематоды принадлежат к быстро эволюционирующей линии (рис. 2), что неизбежно приводит к затруднениям при построении филогенетических деревьев с использованием традиционных методов (Delsuc *et al.*, 2005; Philippe *et al.*, 2005). Это хорошо согласуется с ранее полученными данными (Aguinaldo *et al.*, 1997). Млекопитающие являются наиболее медленно эволюционирующей группой; у вторичноротых (deuterostome) обнаруживается высокий уровень изменчивости по скоростям эволюции, включая быстро эволюционирующих иглокожих и оболочечников (морской еж *St* и *Ciona intestinalis* Ci, рис. 2). Насекомые эволюционируют с меньшей скоростью, чем нематоды; насекомые и нематоды характеризуются относительно небольшой внутригрупповой изменчивостью скоростей эволюции (рис. 2).

К гомоплазии при использовании RGC_CAM метода могут приводить два основных эволюционных события: параллельные изменения (рис. 2) и обратные замены (Nei, Kumar, 2001; Rogozin *et al.*, 2007a, b). Метод RGC_CAM позволяет прямо оценивать возможный уровень гомоплазии (Rogozin *et al.*, 2008). Для получения таких оценок для параллельных замен использовалась схема, представленная на рис. 3. Два или три вида (например, человек–мышь

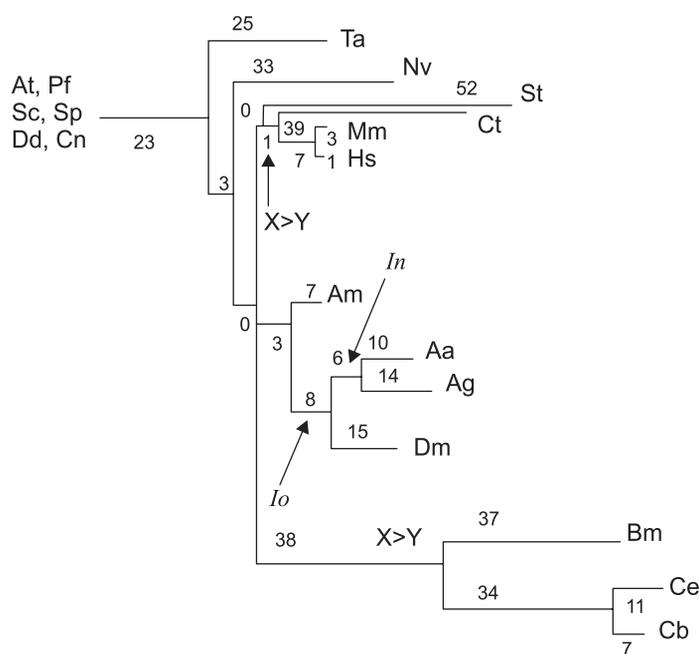


Рис. 2. Филогенетическое дерево, использованное в данном исследовании.

Числа рядом с ветвями дерева соответствуют числу RGC_CAM (длина ветви). Параллельные замены показаны как $X > Y$. *Io* – «старая» внутренняя ветвь насекомых, *In* – «новая» внутренняя ветвь насекомых (см. текст). Сокращения названий видов приведены на рис. 1.

или дрозофила–два вида москитов, рис. 3) характеризуются одной и той же аминокислотой Y, в то время как все остальные виды имеют другую аминокислоту X. В этом случае параллельные изменения являются наиболее вероятным объяснением, поскольку все другие сценарии требуют реализации, по крайней мере, трех событий.

Для метода RGC_CAM важным аспектом является набор видов внешней группы (аутгруппы), относительно которых исследуются аминокислотные замены: аминокислота в исследуемой позиции должна быть консервативна во всех видах аутгруппы и считается предковым состоянием (рис. 1) (Rogozin *et al.*, 2007b, 2008). Нами выполнен анализ комбинаций видов (от 12 до 19 видов), включавших от 1 до 8 внешних групп (аутгрупп) видов: At, Pf, Sc, Sp, Dd, Cn, Ta, Nv (суммарно 255 комбинаций) (рис. 2). Для насекомых и вторичноротых анализировали по две внутренние ветви, одна была ближе к «корню» исследуемого поддерева («старая» ветвь) и другая ближе к «листу» («новая» ветвь) (пример для насекомых приведен на рис. 2 [*Io* и *In*]). В случаях, когда в качестве аутгрупп включались все 8 или комбинации 7 видов,

параллельные RGC_CAM не обнаруживались среди вторичноротых, в ветвях насекомых и нематод наблюдались одно или два параллельных изменения. В отличие от этого для других комбинаций видов в аутгруппе было выявлено значительное количество параллельных изменений между нематодами (внутренняя ветвь *N*) и насекомыми, преимущественно в «старых» внутренних ветвях (табл. 1). «Старая» ветвь у насекомых (*Io*) только ненамного длиннее, чем «новая» ветвь (*In*) (8 RGC_CAM и 6 RGC_CAM соответственно; рис. 2), что не объясняет существенный избыток параллельных изменений в «старых» ветвях по сравнению с «новыми». Такая же тенденция наблюдалась и между «старыми» и «новыми» ветвями насекомых и «новыми» ветвями вторичноротых, со значительно большим количеством параллельных изменений, возникших в «старых» ветвях насекомых (табл. 1). Из этого следует, что скорость параллельных замен неодинакова, и в ряде случаев обнаруживается больше параллельных изменений в более древних ветвях.

Далее мы провели сравнение количества параллельных изменений в «листьях» дерева (ветви, которые ведут к конкретному виду, на-

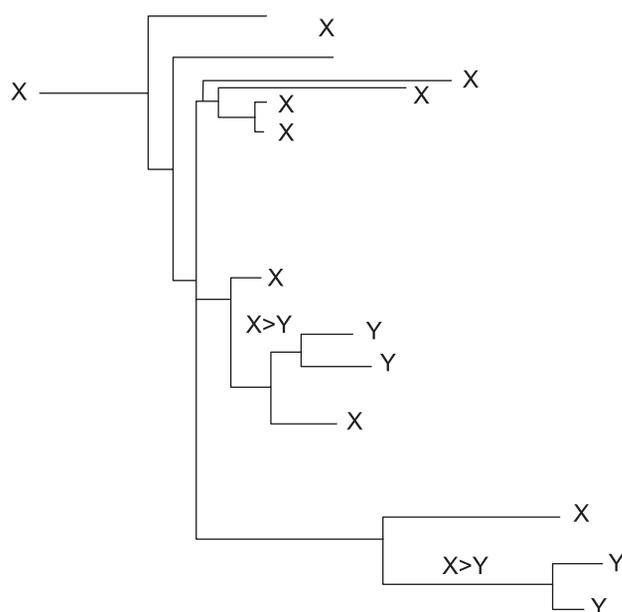


Рис. 3. Идентификация параллельных замен $X > Y$. Филогенетическое дерево аналогично дереву на рис. 2.

пример, ветвь между человеком и последним общим предком человека и мыши). Нами был выполнен сравнительный анализ двух ветвей, ведущих к двум видам москитов и двум видам нематод. «Листья» нематод примерно в 3 раза короче внутренних ветвей нематод, в то время как «листья» москитов примерно в 2 раза длиннее, чем «новые» внутренние ветви у группы видов насекомых (рис. 2). Очевидно, что при их сравнении можно было бы ожидать количество параллельных замен, близкое к выявляемым при сравнении N_{In} (табл. 1). Однако ни в одном из таких 4 сравнений параллельных изменений обнаружено не было. Это не может быть обусловлено малой длиной ветвей, поскольку параллельные замены обнаруживаются при сопоставлении значительно более коротких внутренних «старых» ветвей при сравнении I_o_{Do} (рис. 2). Отсутствие параллельных замен во внутренних «новых» ветвях хорошо согласуется с их повышенным количеством во внутренних «старых» ветвях рассмотренных групп видов, принадлежащих к более удаленным таксонам.

Для того, чтобы оценить статистическую достоверность наблюдаемых тенденций, была применена следующая схема анализа. В результате 255 попарных сравнений отобранных пар ветвей для различных комбинаций аутруп были выбраны уникальные параллельные за-

мены (повторяющиеся параллельные замены были исключены из анализа, поэтому конечное число рассматриваемых замен является объединением полученных в индивидуальных сравнениях). Для N_{I_o} сравнения выявлено 6 уникальных параллельных замен, остальные 3 варианта сравнений позволили выявить только 3 уникальные параллельные замены (табл. 2). Внутренняя ветвь нематод N была одной и той же при всех сравнениях, так что частоты параллельных замен зависели только от длин ветвей внутри группы видов насекомых. Отношение длины ветви I_o к суммарной длине всех других анализируемых ветвей было $\sim 0,2$ (рис. 2, табл. 2). Вероятность случайного выявления в этой короткой ветви 6 из 9 обнаруженных параллельных замен по биномиальному тесту около 0,003.

Несмотря на то что наблюдаемый избыток параллельных замен во внутренних ветвях статистически оказывается высоко достоверным, количество параллельных замен невелико (табл. 2). Для того чтобы увеличить разрешающую способность метода, нами было расширено определение RGC путем включения всех возможных аминокислотных замен (в отличие от соответствующих RGC_CAM, включающих только две или три нуклеотидные замены). Такая расширенная характеристика была обозначена как

Таблица 1

Распределение параллельных замен для нематод, насекомых и вторичноротых для 255 комбинаций видов аутгруппы

Сравниваемые ветви	Число параллельных замен						
	0	1	2	3	4	5	> 5
<i>Io_Do</i>	240	15					
<i>In_Do</i>	255						
<i>Io_Dn</i>	128	127					
<i>In_Dn</i>	255						
<i>N_Io</i>		112	68	56	14	4	1
<i>N_In</i>	192	63					
<i>N_Do</i>	255						
<i>N_Dn</i>	238	16	1				

Обозначения ветвей дерева: N – внутренняя ветвь нематод; Io – «старая» внутренняя ветвь насекомых; In – «новая» внутренняя ветвь насекомых; Do – «старая» внутренняя ветвь вторичноротых; Dn – «новая» внутренняя ветвь вторичноротых (см. текст).

Таблица 2

Сравнение параллельных замен во внутренних и внешних ветвях нематод, насекомых и вторичноротых

Параметр	Исследуемые ветви дерева			
	<i>N_Io</i>	<i>N_In</i>	<i>N_Ag</i>	<i>N_Aa</i>
Число RGC_CAM, %	6 (67)	1 (11)	0 (0)	2 (22)
Относительная длина ветви	0,201	0,169	0,365	0,265
Нормализованное число параллельных замен	29,9	5,9	0	7,5
Число RGC_CA, %	45 (45)	22 (22)	15 (15)	19 (19)
Относительная длина ветви	0,205	0,205	0,340	0,250
Нормализованное число параллельных замен	219,5	107,3	44,1	76
	<i>N_Do</i>	<i>N_Dn</i>	<i>N_Hs</i>	<i>N_Mm</i>
Число RGC_CAM, %	0 (0)	2 (100)	0 (0)	0 (0)
Относительная длина ветви	0,094	0,621	0,109	0,176
Нормализованное число параллельных замен	0	3,2	0	0
Число RGC_CA, %	28 (33)	51 (60)	3 (3,5)	3 (3,5)
Относительная длина ветви	0,121	0,581	0,139	0,159
Нормализованное число параллельных замен	231,4	82,1	21,6	18,9
	<i>Io_Do</i>	<i>Io_Dn</i>	<i>Io_Hs</i>	<i>Io_Mm</i>
Число RGC_CAM, %	1 (50)	1 (50)	0 (0)	0 (0)
Относительная длина ветви	0,094	0,621	0,109	0,176
Нормализованное число параллельных замен	10,6	1,6	0	0
Число RGC_CA, %	21 (45)	24 (51)	1 (2)	1 (2)
Относительная длина ветви	0,121	0,581	0,139	0,159
Нормализованное число параллельных замен	173,6	41,3	7,2	6,3
	<i>In_Do</i>	<i>In_Dn</i>	<i>In_Hs</i>	<i>In_Mm</i>
Число RGC_CAM, %	0 (0)	0 (0)	0 (0)	0 (0)
Относительная длина ветви	0,094	0,621	0,109	0,176
Число RGC_CA, %	5 (23)	16 (72)	0 (0)	1 (5)
Относительная длина ветви	0,121	0,581	0,139	0,159
Нормализованное число параллельных замен	41,3	27,5	0	6,3

Обозначения ветвей дерева: N – внутренняя ветвь нематод; Io – «старая» внутренняя ветвь насекомых; In – «новая» внутренняя ветвь насекомых; Do – «старая» внутренняя ветвь вторичноротых; Dn – «новая» внутренняя ветвь вторичноротых (см. текст); Hs – ветвь, ведущая к человеку; Mm – ветвь, ведущая к мышши; Aa – ветвь, ведущая к москиту *Aedes aegypti*; Ag – ветвь, ведущая к москиту *Anopheles gambiae*. Относительная длина ветви – это отношение длины исследуемой ветви к общей длине ветвей, усредненное для 255 комбинаций аутгрупп. Нормализованное число параллельных замен – это число параллельных замен, деленное на относительную длину ветви.

RGC_CA. Ослабление исходного определения RGC_CAM должно было бы привести к существенному увеличению числа гомоплазии: должно наблюдаться большее количество параллельных изменений. В то же время их увеличение могло быть обусловлено включением в анализ более чем двух параллельных и/или обратных замен. Хотя численно избыток параллельных замен при сравнении во внутренних ветвях был менее существенным, чем при использовании метода RGC_CAM, статистическая обоснованность наблюдаемых отличий оказалась даже еще более выраженной ($P < 10^{-7}$) благодаря большому количеству наблюдений (табл. 2). Такая же тенденция наблюдалась и при сравнении других коротких внутренних ветвей, таких, как *Do* и *Dn*, когда параллельные замены оценивались по методу RGC_CA (табл. 2). Это указывает на то, что выявленная тенденция не зависит от метода анализа.

Одним из ожидаемых результатов выявленной тенденции повышенного количества параллельных замен во внутренних «старых» ветвях рассмотренных групп видов, принадлежащих к более удаленным таксонам, является более высокая частота параллельных замен для более близких таксонов. Для того чтобы исследовать этот вопрос, был проведен анализ внутри таксона насекомых (табл. 3). Число RGC_CAM мало и не поддается интерпретации (табл. 3). Однако RGC_CA ведут себя в соответствии с

выявленной тенденцией: число RGC_CA для сравнения *Am_In* (62, табл. 3) в 4 раза выше числа RGC_CA для сравнения *In_Dn* (16, табл. 2), несмотря на то что длины ветвей примерно одинаковы (рис. 2). Это хорошо соответствует предсказанию, что параллельные замены происходят с повышенной частотой между более близкими таксонами.

Обсуждение

Одной из задач настоящей работы было выяснение возможности прямой оценки скорости возникновения параллельных изменений при использовании метода RGC_CAM в филогенетических исследованиях. В процессе выполнения такого анализа были обнаружены отличия в распределении параллельных изменений между ветвями филогенетических деревьев с их избытком между наиболее «старыми» ветвями по сравнению с более «новыми». Такое отсутствие единообразия параллельных изменений в зависимости от «древности» сравниваемых ветвей создает дополнительную проблему при филогенетическом анализе с использованием RGC_CAM и других методов.

Гомоплазия является одной из основных проблем при использовании филогенетических методов анализа, в то же время это один из аспектов молекулярной эволюции, который заслуживает специального внимания. В частности, данные

Таблица 3
Сравнение параллельных замен во внутренних и внешних ветвях насекомых

Параметр	Исследуемые ветви дерева		
	<i>Am_In</i>	<i>Am_Ag</i>	<i>Am_Aa</i>
Число RGC_CAM, %	2 (25)	4 (50)	2 (25)
Относительная длина ветви	0,211	0,457	0,332
Нормализованное число параллельных замен	9,5	8,6	6,0
Число RGC_CA, %	62 (54)	37 (33)	15 (13)
Относительная длина ветви	0,258	0,428	0,315
Нормализованное число параллельных замен	240,3	86,4	47,6

Обозначения ветвей дерева: *Am* – ветвь, ведущая к пчеле; *Aa* – ветвь, ведущая к москиту *Aedes aegypti*; *Ag* – ветвь, ведущая к москиту *Anopheles gambiae*. Исследовалось три варианта сравнений ветвей: 1) *Am_In*: *Am* = Y, *Dm* = X, *Ag* = Y, *Aa* = Y, все остальные виды = X; 2) *Am_Ag*: *Am* = Y, *Dm* = X, *Ag* = Y, *Aa* = X, все остальные виды = X; 3) *Am_Aa*: *Am* = Y, *Dm* = X, *Ag* = X, *Aa* = Y, все остальные виды = X. Относительная длина ветви – это отношение длины исследуемой ветви к общей длине ветвей, усредненное для 255 комбинаций аутгрупп. Нормализованное число параллельных замен – это число параллельных замен, деленное на относительную длину ветви.

о выраженном избытке параллельных замен во внутренних ветвях филогенетических деревьев по сравнению с более «новыми» ветвями нуждается в объяснении. Оно может быть найдено в коварионной модели молекулярной эволюции (Fitch, Markowitz, 1970; Fitch, 1976; Penny *et al.*, 2001). Согласно этой модели, в каждый данный момент эволюции только для небольшого (относительно) количества аминокислотных позиций (обозначаемых как коварионы: одновременно переменные кодоны – **concomitantly variable codons**) в белках разрешены замены, тогда как замены в других позициях не фиксируются в связи с функциональными ограничениями. Фиксировавшиеся замены изменяют совокупность коварионов, приводя к появлению новых коварионов и разрушению существующих коварионов. В течение длительных эволюционных периодов возникающие «волны» коварионов занимают всю или большую часть последовательности белка. В этой модели эволюции две расходящиеся линии имеют один и тот же набор коварионов, но сразу же после дивергенции со временем, постепенно, перекрывание между набором коварионов уменьшается. Это происходит из-за того, что вероятность параллельных изменений последовательно уменьшается вместе с уменьшением сходства рядов коварионов в расходящихся линиях. Необходимо отметить, что это не исключает появления параллельных замен у сильно дивергировавших линий, поскольку одни и те же аминокислотные позиции в процессе эволюции могут вовлекаться в разные наборы коварионов у расходящихся линий.

В более широком эволюционном контексте наблюдаемый избыток параллельных замен, по-видимому, соответствует одному из механизмов, лежащих в основе известного закона «гомологических рядов в наследственной изменчивости». Закон гомологических рядов был сформулирован выдающимся русским генетиком, ботаником и геоботаником Н.И. Вавиловым как обобщение результатов сбора данных и их анализа об изменчивости фенотипических признаков у культурных и диких видов растений (Vavilov, 1922). Первое положение этого закона имеет следующую формулировку: «Виды и роды, генетически близкие, характеризуются сходными рядами наследственной изменчивости с такой правильностью, что, зная ряд форм в пределах

одного вида, можно предвидеть нахождение параллельных форм у других видов и родов. Чем ближе генетически расположены в общей системе роды и линнеоны (виды), тем полнее сходство в рядах их изменчивости» (Vavilov, 1922). Исходя из прогностических возможностей в отношении параллельной изменчивости у разных таксонов, этот закон был воспринят как «таблица Менделеева в биологии» (Hammer, Schubert, 1994; Popov, 2002; Zakharov, 2005; Bogdanov *et al.*, 2007; Юрченко, Захаров, 2007; Шумный, 2007), однако до сих пор остаются недостаточно исследованными возможные молекулярно-генетические основы реализации такого сходства (Hammer, Schubert, 1994). Следует подчеркнуть, что несмотря на почти вековую историю его открытия и огромное количество подтверждающих его наблюдений, этот закон до сих пор остается на уровне описания феномена, механизмы которого неизвестны. С нашей точки зрения, избыток параллельных изменений в глубоких внутренних («старых») ветвях является одной из молекулярных основ закона Н.И. Вавилова.

Заключение

Избыток параллельных изменений в глубоких внутренних («старых») ветвях сам по себе является заметным эволюционным феноменом. Логичным объяснением этого феномена является коварионная модель молекулярной эволюции. Важно подчеркнуть, что такая коварионная преддетерминированность событий может лежать и в основе одного из возможных молекулярно-генетических механизмов реализации закона о рядах гомологической изменчивости Н.И. Вавилова. Полученные результаты свидетельствуют о том, что сам феномен параллелизма в изменчивости может иметь молекулярно-генетическую природу, лежащую в ограничении спектра возможных замен имеющимися рядами коварионов. То есть параллелизм в изменчивости, совпадающий с близостью происхождения, может быть обусловлен, прежде всего, наличием общих механизмов «запрета» на фиксацию замен, которые определяются характеристиками предсуществующих совокупностей коварионов. Интересно, что эта точка зрения совпадает с трактовкой Ч. Дарвина параллельной изменчивости как случаев, зависящих от неизвестных причин, действующих на

одинаковые конституции, а отсюда и варьирующие одинаково (Darwin, 1859), если под одинаковой конституцией понимать совокупность коварионов.

Литература

- Вавилов Н.И. Закон гомологических рядов в наследственной изменчивости. 2-е изд., перераб. и расш. М.: Сельхозгиз, 1935. 56 с.
- Филиппченко Ю.А. О параллелизме в живой природе // Усп. эксперим. биологии. 1924. Т. 3. Вып. 3/4. С. 248.
- Шумный В.К. Два гениальных обобщения Николая вановича Вавилова (к 120-летию со дня рождения) // Генетика. 2007. Т. 43. № 11. С. 1447–1584.
- Юрченко Н.Н., Захаров И.К. Концепция биологической гомологии: исторический обзор и современные взгляды // Информ. вестник ВОГиС. 2007. Т. 11. № 3/4. С. 537–546.
- Aguinaldo A.M., Turbeville J.M., Linford L.S. *et al.* Evidence for a clade of nematodes, arthropods and other moulting animals // Nature. 1997. V. 387. № 6632. P. 489–493.
- Averof M., Rokas A., Wolfe K.H., Sharp P.M. Evidence for a high frequency of simultaneous double-nucleotide substitutions // Science. 2000. V. 287. № 5456. P. 1283–1286.
- Bogdanov Y.F., Grishaeva T.M., Dadashev S.Y. Similarity of the domain structure of proteins as a basis for the conservation of meiosis // Int. Rev. Cytol. 2007. V. 257. P. 83–142.
- Boore J.L. The use of genome-level characters for phylogenetic reconstruction // Trends Ecol. Evol. 2006. V. 21. № 8. P. 439–446.
- Darwin C. On the Origin of Species by Means of Natural Selection: or the Preservation of Favoured Races in the Struggle for Life. London: John Murray, 1859. (1-st ed.).
- Delsuc F., Brinkmann H., Philippe H. Phylogenomics and the reconstruction of the tree of life // Nat. Rev. Genet. 2005. V. 6. № 5. P. 361–375.
- Fitch W.M. The molecular evolution of cytochrome c in eukaryotes // J. Mol. Evol. 1976. V. 8. № 1. P. 13–40.
- Fitch W.M., Markowitz E. An improved method for determining codon variability in a gene and its application to the rate of fixation of mutations in evolution // Biochem. Genet. 1970. V. 4. № 5. P. 579–593.
- Hammer K., Schubert I. Are Vavilov's law of homologous series and synteny related? // Genet. Res. and Crop Evol. 1994. V. 41. P. 123–124.
- Hennig W. Grundzüge einer Theorie der Phylogenetischen Systematik. Berlin: Deutscher Zentralverlag, 1950.
- Nei M., Kumar S. Molecular Evolution and Phylogenetics. Oxford: Oxford Univ., 2001.
- Penny D., McComish B.J., Charleston M.A., Hendy M.D. Mathematical elegance with biochemical realism: the covarion model of molecular evolution // J. Mol. Evol. 2001. V. 53. № 6. P. 711–723.
- Philippe H., Zhou Y., Brinkmann H. *et al.* Heterotachy and long-branch attraction in phylogenetics // BMC Evol. Biol. 2005. V. 5. P. 50.
- Popov I.Y. Periodical systems in biology // Die Entstehung biologischer Disziplinen / Ed. U. Hossfeld, T. Junker. Berlin: VWB, 2002. P. 55–69.
- Rogozin I.B., Thompson K., Csuros M. *et al.* Homoplasy in genome-wide analysis of rare amino acid replacements: the molecular-evolutionary basis for Vavilov's law of homologous series // Biol. Direct. 2008. V. 3. P. 7.
- Rogozin I.B., Wolf Y.I., Carmel L., Koonin E.V. Ecdysozoan clade rejected by genome-wide analysis of rare amino acid replacements // Mol. Biol. Evol. 2007a. V. 24. № 4. P. 1080–1090.
- Rogozin I.B., Wolf Y.I., Carmel L., Koonin E.V. Analysis of rare amino acid replacements supports the Coelomata clade // Mol. Biol. Evol. 2007b. V. 24. № 12. P. 2594–2597.
- Rokas A., Holland P.W. Rare genomic changes as a tool for phylogenetics // Trends Ecol. Evol. 2000. V. 15. № 11. P. 454–459.
- Silva J.C., Kondrashov A.S. Patterns in spontaneous mutation revealed by human-baboon sequence comparison // Trends Genet. 2002. V. 18. № 11. P. 544–547.
- Sinha H., Nicholson B.P., Steinmetz L.M., McCusker J.H. Complex genetic interactions in a quantitative trait locus // PLoS Genetics. 2006. V. 2. № 2. P. 140–147.
- Snel B., Huynen M.A., Dutilh B.E. Genome trees and the nature of genome evolution // Annu. Rev. Microbiol. 2005. V. 59. P. 191–209.
- Tatusov R.L., Fedorova N.D., Jackson J.D. *et al.* The COG database: an updated version includes eukaryotes // BMC Bioinformatics. 2003. V. 4. P. 41.
- Telford M.J., Budd G.E. The place of phylogeny and cladistics in Evo-Devo research // Int. J. Dev. Biol. 2003. V. 47. № 7/8. P. 479–490.
- Wolf Y.I., Rogozin I.B., Grishin N.V., Koonin E.V. Genome trees and the tree of life // Trends Genet. 2002. V. 18. № 9. P. 472–479.
- Vavilov N.I. The law of homologous series in variation // J. Genet. 1922. V. 12. P. 47–89.
- Zakharov I.A., Nikolai I. Vavilov (1887–1943) // J. Biosci. 2005. V. 30. № 3. P. 299–301.

THE MOLECULAR-EVOLUTIONARY BASIS FOR VAVILOV'S LAW OF HOMOLOGOUS SERIES

I.B. Rogozin^{1,2}, V.I. Glazko³, E.V. Koonin¹

¹National Center for Biotechnology Information NLM, National Institutes of Health, Bethesda, MD 20894, USA, e-mail: rogozin@mail.nih.gov; ²Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia; ³Timiriazev's State Agricultural University, Moscow, Russia

Summary

Rare genomic changes (RGCs) that are thought to comprise derived shared characters of individual clades are becoming an increasingly important class of markers in genome-wide phylogenetic studies. Recently, we proposed a new type of RGCs designated RGC_CAMs (after Conserved Amino acids-Multiple substitutions) that were inferred using genome-wide identification of amino acid replacements that were: i) located in unambiguously aligned regions of orthologous genes, ii) shared by two or more taxa in positions that contain a different, conserved amino acid in a much broader range of taxa, and iii) require two or three nucleotide substitutions. We provide a direct estimate of the level of homoplasy caused by parallel changes, one of the major classes of events leading to homoplasy, among the RGC_CAMs using 462 alignments of orthologous genes from 19 eukaryotic species. Parallel changes occur much more often in relatively recently diverged lineages than in those separated from their last common ancestor by longer time intervals of time. This pattern seems to provide the molecular-evolutionary underpinning of Vavilov's law of homologous series and is readily interpreted within the framework of the covarion model of molecular evolution.