

УДК 575.852.112:575.852.113:004.273

К.В. Гунбин¹, М.А. Генаев¹, И.И. Турнаев¹, Д.А. Афонников^{1,2}

¹Институт цитологии и генетики СО РАН (г. Новосибирск)

²Новосибирский государственный университет (г. Новосибирск)

КОМПЬЮТЕРНАЯ СИСТЕМА АНАЛИЗА РЕЖИМОВ МОЛЕКУЛЯРНОЙ ЭВОЛЮЦИИ ГЕНОВ И БЕЛКОВ: АНАЛИЗ ЭВОЛЮЦИИ ЦИКЛИНОВ В

Работа выполнена при финансовой поддержке РФФИ (грант № 09-04-01641-а), а также в рамках интеграционных проектов СО РАН № 113, 119; Министерства образования и науки РФ (ГК № П857, № 07.514.11.4003); Программ РАН (№ 6.8, Б 26.29 и 24.2).

Создана интернет-доступная компьютерная система анализа режимов молекулярной эволюции генов и белков (<http://pixie.bionet.nsc.ru/satem/>). Анализ режимов эволюции производится на основе расчета отношения частот фиксации радикальных аминокислотных замен к консервативным, на основе впервые предложенного нами метода анализа скоростей различных типов замен аминокислот в эволюции белков и на основе статистического соотнесения изменения свойств аминокислот с фенотипическими признаками организмов. С помощью созданной системы проведено исследование режимов молекулярной эволюции циклинов В. Показано качественное отличие картин фиксации атипичных аминокислотных замен в эволюции циклинов В животных и грибов. Также показано, что у животных эволюция циклинов В3, в отличие от циклинов В1 и В2, была связана с усложнением организма.

Ключевые слова: молекулярная эволюция; консервативные и радикальные аминокислотные замены; система «клиент-сервер»; статистический анализ; циклины В.

Введение

В последнее время появилось огромное количество данных о существовании универсальной связи между уровнем экспрессии гена и скоростью его эволюции [1–3]. В 2008 г. впервые было показано, что эта связь обусловлена отбором против неправильного сворачивания белков вследствие ошибок трансляции [2]. Также показано [2], что этот отбор может проявляться в: 1) уменьшении числа ошибок трансляции, 2) уменьшении вероятности неправильного сворачивания белка при ошибочной трансляции, 3) уменьшении вероятности неправильного сворачивания и денатурации белка. Очевидно, что давление отбора в результате ответа на процессы (2) и (3) связаны с аминокислотными заменами, по-разному влияющими на структуру белка. Структура белка, его функция и сворачивание определяются комбинацией свойств аминокислот. Проанализировано более 500 свойств аминокислот

[4], что позволяет проводить систематические исследования эволюции белок-кодирующих генов. Подходами, направленными на решение этой задачи, являются анализ отношения частот фиксации радикальных аминокислотных замен к консервативным (K_R/K_C) [5, 6] и исследование скоростей изменения физико-химических свойств белков (V_{PC}) в их эволюции [7]. Однако эти подходы обладают существенным недостатком – необходимо знать заранее «адаптивность изменения» определенного свойства аминокислот. Предложенный нами подход базируется на анализе K_R/K_C и V_{PC} , но имеет два отличия: 1) при анализе используются все известные свойства аминокислот; 2) проводится статистическое соотнесение изменения этих свойств с фенотипическими признаками организмов, что позволяет напрямую связать молекулярную эволюцию с приспособительной эволюцией организмов. На этой основе нами была создана компьютерная система для анализа режимов молекулярной эволюции генов и белков (<http://pixie.bionet.nsc.ru/samem/>).

В настоящей работе эта система использована для анализа эволюции циклинов В. Известно, что циклины В составляют древнее по происхождению семейство белков, контролирующих центральную часть клеточного цикла – вход в митоз [8, 9]. Важно отметить, что только эти циклины экспрессируются во всех тканях организма животных [8, 9]. В то же время детальный анализ режимов молекулярной эволюции циклинов В до сих пор не проведен. Поэтому с целью тестирования возможностей созданной компьютерной системы был проведен анализ эволюции циклинов В.

Материалы и методики исследования

Описание компьютерной системы

Система состоит из двух основных конвейеров, анализа эволюции генов и анализа эволюции белков, и двух дополнительных, собирающих выборки генов и белков и производящих их первичный анализ, построение матриц BLOSUM. Основные конвейеры позволяют провести основные этапы обработки данных, комбинируя различные методы множественного выравнивания, построения филогенетического дерева и реконструкции предковых последовательностей. Множественное выравнивание рассчитывается программами MAFFT 6.717 и KALIGN 2.04; филограмма (филогенетическое дерево с неравными длинами ветвей от корня, отражающими скорости замен) – программами FASTTREE 2.1.1 и PHYML 3.0; филограмма преобразуется в хронограмму (филогенетическое дерево, построенное на основе гипотезы релаксированных молекулярных часов) программой R8S 1.71. При преобразовании филограммы в хронограмму используются датировки дивергенций, задаваемые пользователем. Предковые последовательности белков реконструируются на основе выравниваний, не содержащих делеции, с помощью программ ANCESCON, FASTML (серверная версия) и CODEML

(из пакета PAML 4.4); генов – ANC-GENE, FASTML (серверная версия) и CODEML (из пакета PAML 4.4). Пользователь может рассчитать модель эволюции заданного семейства белков, используя MODELESTIMATOR 1.1, или воспользоваться обобщенными моделями. Для анализа используются данные о 531 свойстве аминокислот [4].

Общий модуль оценки взаимосвязи эволюции свойств аминокислот и фенотипических признаков реализован на языке R (пакет Ape) и реализует 3 группы статистических методов анализа данных [10] (GEE – Generalized Estimating Equations, Lynch, или метод Variance Partitioning, и GLS – Generalized Least Squares), принимающих во внимание филогенетическую инерцию (наведенную хронограммой корреляцию [10]). Все использованные программы снабжены гиперссылками на соответствующий литературный источник, что позволяет пользователю своевременно знакомиться с методологической базой каждой используемой им программы. Для облегчения работы пользователя для каждого вычислительного этапа компьютерной системы имеются образцы вводимых данных.

Существенной особенностью конвейера анализа генов является возможность исследования разными методами оценки K_R/K_C . Пользователь может воспользоваться как методами Жанга (программа HON-NEW) [5] и Смита [6], так и улучшенным нами методом Смита. При анализе данных методами Жанга и Смита пользователь может задать число групп для разделения 20 аминокислот. Для каждого из 531 свойств аминокислоты группируются методом К-средних (используя R). В улучшенном нами методе Смита [6] для разбиения аминокислотных замен на консервативные и радикальные используется информация из матрицы BLOSUM, строящейся для каждого анализируемого семейства белков. Все неотрицательные (≥ 0) значения в матрице BLOSUM трактуются нами как консервативные, все отрицательные – как радикальные.

Таким образом, в отличие от ранее предложенных подходов [5, 6], в которых радикальность и консервативность аминокислотных замен задается исходя из группировки аминокислот по заранее заданным физико-химическим свойствам, нами используется более естественное разбиение аминокислотных замен, учитывающее структурную и функциональную природу исследуемых белков. Важно отметить, что на отношение K_R/K_C существенно влияют композиционные характеристики исследуемых генов, такие как частоты использования кодонов и отношение транзиций к трансверсиям [6]. Для сравнения наблюдаемых значений K_R/K_C со значениями K_R/K_C , свойственными эволюции генов при $K_A/K_S=1$, с помощью пакета INDELible 1.03 пользователем может быть проведено Марковское моделирование эволюции генов, учитывающее естественные частоты использования кодонов и естественное отношение транзиций к трансверсиям. Следует отметить, что в конвейере анализа эволюции генов также реализована возможность суммирования всех рассчитываемых величин (K_R/K_C , K_C и K_R).

При анализе последовательностей белков существенной особенностью является впервые предложенный нами метод анализа скоростей фиксации различных типов аминокислотных замен. Он основан на Марковском моделировании эволюции белков пакетом INDELible 1.03 и непараметрическом перестановочном тесте, предложенном нами в работе [11]. INDELible 1.03 позволяет учесть детальные особенности эволюции белков (различие частот встречаемости аминокислот и частот их замен у паралогичных групп внутри одного семейства белков вплоть до учета особенностей эволюции на отдельных ветвях дерева). Тест с перестановками [11] позволяет выявить для каждой ветви филогенетического дерева характерные типы аминокислотных замен (атипичные замены), частота которых статистически значимо превышает частоту в смоделированных последовательностях.

Метод состоит из трех этапов анализа. На первом этапе на каждой ветви дерева подсчитывается наблюдаемое количество ($n_{real(Type)}$) всех 190 типов аминокислотных замен. Для внутренних ветвей это делается путем парного сравнения последовательности-предка и последовательности-потомка, реконструированных программами CODEML, FASTML или ANCESCON. Для конечных ветвей наблюдаемые числа всех типов аминокислотных замен ($n_{real(Type)}$) подсчитываются путем парного сравнения последовательности реконструированного предка и последовательности-потомка из современного организма. На втором этапе на каждой ветви дерева подсчитываются ожидаемые количества всех типов замен на основе заданного пользователем количества компьютерных симуляций молекулярной эволюции белков. Для этого анализируются более 1 000 модельных выравниваний, генерируемых программой INDELible 1.03 на основе заданных пользователем особенностей эволюции белков. На третьем этапе для каждой ветви дерева с помощью перестановочного теста идет сравнение числа ожидаемых и наблюдаемых замен каждого из 190 типов. Множество наблюдаемых замен при достаточном числе симуляций всегда входит во множество ожидаемых замен.

Следовательно, путем случайной перестановки числа наблюдаемых замен внутри модельного набора замен можно сгенерировать заданное пользователем число ($> 1\,000$) случайных выборок замен N , равных по объему выборкам наблюдаемых замен. Для каждой такой случайной выборки в тесте оценивается число $n_{rand(Type)}$ аминокислотных замен определенного типа. Далее подсчитывается число случайных выборок M , в которых $n_{rand(Type)} > n_{real(Type)}$. Величина M/N оценивает вероятность p , с которой встречаемость аминокислотных замен определенного типа, наблюдаемая в исходной выборке генов, могла возникнуть по случайным причинам. Следует отметить, что в конвейере анализа эволюции белков также реализована возможность суммарного анализа эволюционных изменений физико-химических свойств аминокислот. Для этого реализованы коррекции эволюционных расстояний Кимуры, Джукса–Кантора и Гамма-коррекция.

Техническая основа компьютерной системы

Компьютерная система анализа режимов молекулярной эволюции генов и белков создана на основе клиент-серверной конвейерной системы обработки данных BioinfoWF. Серверная часть системы реализована в виде приложения на языке Perl и отвечает за запуск и выполнение конвейера. На вход приложению подаются схема конвейера и описание вычислительных модулей. Схема конвейера, форматы данных и взаимодействие вычислительных модулей в ней описываются на языке XML. Описание конвейера состоит из двух файлов. Первый описывает вычислительные модули, второй задает топологию конвейера. Описание модулей состоит из следующих разделов: 1) входные файлы; 2) выходные файлы; 3) параметры и опции (идентификатор, описание, тип параметра, значение по умолчанию, внешний вид поля запроса значения для параметра на странице web-браузера); 4) правила генерации командной строки, 5) правила действия пользовательского интерфейса, позволяющие динамически изменять web-интерфейс модуля в процессе редактирования конвейера. Второй файл задает топологию конвейера. В нем описываются имена входных/выходных файлов для каждого модуля и значения для каждого параметра или опции. Приложение запускает конвейер, создавая файл с отчетом о статусе выполнения каждого узла в формате XML. Серверная часть поддерживает параллельный запуск узлов конвейера и позволяет производить удаленный запуск ресурсоемких расчетных модулей на вычислительном кластере. Режим запуска узла (локальный или удаленный) задается в схеме конвейера.

Клиентская часть реализована в виде web-приложения. Пользователю предлагается работать с уже готовыми схемами. При этом доступны базовые опции управления конвейером: 1) установка входных файлов, параметров и опций для каждого вычислительного модуля в конвейере; 2) старт с произвольного узла и остановка на произвольном узле в конвейере; 3) отслеживание статуса выполнения каждого вычислительного модуля в конвейере; 4) просмотр входных/выходных файлов для каждого этапа расчета в конвейере; 5) привязка форматов входных/выходных файлов к различным приложениям для их визуализации. Генерация страниц в формате HTML реализована на Perl с использованием библиотеки HTML::Template. Дополнительно используется JavaScript библиотека jQuery.

Материалы для тестирования компьютерной системы

В работе использовались 98 белковых последовательностей циклинов В из полностью секвенированных геномов животных, грибов и простейших, представленных в базе данных KEGG 52.0 [12]. Циклины – эволюционно консервативная группа белков, поэтому дерево, построенное по молекулярным данным об их последовательностях, содержало полифуркации, что не всегда позволяло достоверно установить порядок дивергенции некоторых

таксонов. В этом случае порядок дивергенции разрешался вручную на основе анализа литературных данных и интернет-ресурса Tree of Life [13]. Для определения порядка дивергенции одноклеточных предков многоклеточных животных и грибов использовались полногеномные данные, полученные в работах [14–17]. Филогенетические отношения внутри таксона грибов разрешались также на основе полногеномных данных, полученных из работ [18, 19]. Для установления топологии дерева на уровне расхождения нематод, артропод и вторичноротых использовались данные, свидетельствующие в пользу существования группы Ecdysozoa [20–27]. Для разрешения дерева на уровне расхождения иглокожих, головохордовых, хордовых и оболочников использовались данные работ [22, 28–30]. Порядок расхождения организмов внутри класса млекопитающих верифицировался по данным работ [31–35], внутри группы артропод – по данным работы [36]. Порядок дивергенции паралога циклинов грибов, группы организмов, наиболее подверженной эффектам гетеротаксии в эволюции [37], согласовывался с данными работы [38].

В настоящее время оценки числа клеточных типов, являющиеся устойчивой мерой сложности эукариот, известны не для всех организмов, геномные данные которых были взяты нами для анализа. Однако ранее Вогелем и Чотией было проведено исследование взаимосвязи между дубликациями генов различных суперсемейств из базы данных Superfamily 1.69 с количеством клеточных типов для 38 видов эукариотических организмов [39]. Было выявлено 65 суперсемейств, из 1221 у которых число генов семейства в геноме значимо (при $r \geq 0,9$) коррелирует с числом клеточных типов организма (SCOP ID: 48726, 69179, 56854, 47986, 100895, 82671, 57440, 50353, 57196, 57535, 57552, 47090, 57581, 63763, 57501, 57667, 54511, 57256, 63748, 63501, 47769, 48092, 57610, 57302, 69687, 111418, 55550, 47031, 101494, 49842, 50923, 50370, 82708, 110035, 49299, 109775, 48670, 50044, 57586, 47686, 49265, 69318, 82895, 47391, 46966, 47655, 50729, 82615, 47719, 103359, 103637, 54919, 51045, 81783, 48305, 47454, 81419, 57190, 48097, 50242, 50814, 48619, 103575, 90188, 57603). Мы предположили, что выявленные зависимости можно использовать и для оценки числа клеточных типов организмов, не представленных в выборке Вогеля и Чотия. Поэтому в качестве параметра, характеризующего сложность организма, мы использовали десятичный логарифм суммы числа белков в этих суперсемействах (по данным базы Superfamily 1.73 [40]). Чем больше значение этого параметра, тем выше сложность исследуемого организма.

После видового названия приведены идентификаторы последовательностей в базе KEGG. Оцененная при расчете филогенетического дерева достоверность ветвления показана числовыми значениями (критерий Шимодары и Хасегавы / критерий χ^2). Бифуркации, разрешенные вручную на основании литературных данных, показаны звездочкой, по эволюции циклинов грибов – решеткой. Жирным выделены ветви, на которых наблюдаются атипичные замены ($p < 0,01$).

Результаты исследования и обсуждение

В целях тестирования возможностей компьютерной системы анализа режимов молекулярной эволюции генов и белков, а также тестирования разрешающей способности впервые предложенного нами метода анализа скоростей фиксации различных типов аминокислотных замен был проведен анализ эволюции циклинов В животных, грибов и простейших.

На первом этапе анализа мы провели поиск атипичных аминокислотных замен на каждой ветви дерева циклинов В. Для поиска атипичных аминокислотных замен нами были использованы все три программы реконструкции предковых последовательностей белков во внутренних узлах древа циклинов В (при реконструкции с помощью программы ANCESCON учитывалась обобщенная WAG-модель относительных скоростей замен аминокислот; с помощью FASTML – обобщенная LG-модель относительных скоростей замен аминокислот; с помощью CODEML – модель относительных скоростей замен аминокислот для семейства циклинов В, созданная с использованием MODELESTIMATOR), установленные в созданную нами компьютерную систему. На рис. 1 жирной линией показаны ветви, на которых были идентифицированы события атипичных аминокислотных замен с вероятностью $p < 0,01$ при анализе данных всеми тремя методами реконструкции предков.

Рассматривая эволюцию циклинов В животных (рис. 1), можно отметить важную закономерность: у позвоночных накопление атипичных аминокислотных замен, свидетельствующее об ускорении эволюции, идет преимущественно в ходе или после дупликаций паралога. Одна из моделей эволюции функций генов посредством их дупликации предполагает наличие нескольких функций у гена [41]. Очевидно, что после дупликации многофункционального гена эволюция разных функций гена-предка может идти самостоятельно в дублированных копиях гена [41], что увеличивает адаптивную пластичность организма. Интересно, что именно такая картина свойственна эволюции паралогичных групп циклинов В1 и В2 животных, что согласуется с отличиями функций циклинов В1 и В2 животных. Например, известно, что циклин В1 является основным, а циклин В2 – второстепенным, так как мышата с нуль-мутацией по циклину В1 не рождаются, а мышата с нуль-мутацией по циклину В2 рождаются практически нормальными [42]. Также известно, что циклин В1 экспрессируется в интерфазе в двух фракциях: свободной цитоплазматической и связанной с мембраной, а циклин В2 экспрессируется только в мембрано-связанной форме, поэтому циклин В1 практически полностью компенсирует отсутствие циклина В2, но не наоборот [42].

Несколько иная картина свойственна эволюции циклинов В грибов (см. рис. 1): повышение темпов накопления атипичных замен характерно только для одного из дублированных циклинов – сначала только для CLB5-6, затем только для CLB1-2 и, наконец, для отдельной подгруппы циклинов внутри паралогичной группы CDC13.

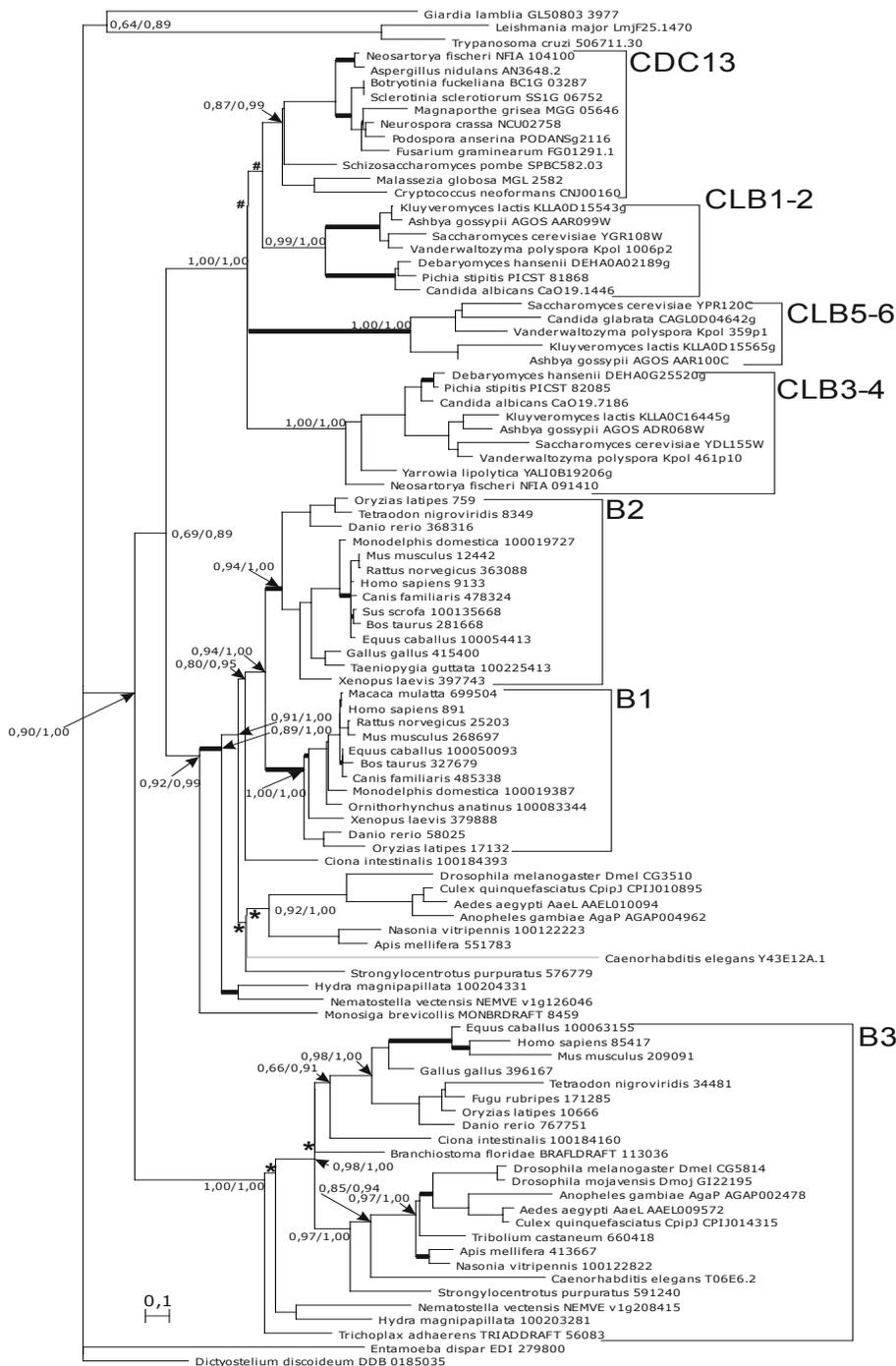


Рис. 1. Филогенетическое дерево циклинов B

Согласно модели дубликаций С. Оно [43] за процессом дубликации гена, осуществляющего одну функцию, следует движущий отбор в одной из копий гена, что приводит к драматическому изменению ее функции. В подавляющем большинстве случаев одна из копий гена в конечном итоге утрачивает свою функцию за счет дегенеративных мутаций. Однако иногда обе копии гена сохраняются в результате случайного приобретения новой важной функции одной из копий гена. Ранее было показано, что для дрожжей имеет место повышенная селекция против дублицированных копий генов; дублицированные же копии генов, сохранившиеся в процессе эволюции, имеют функцию, отличную от предковой [44]. Согласно данным, полученным в настоящей работе, этот вывод можно экстраполировать на эволюцию циклинов В грибов, так как картина распределения атипичных замен на филогенетическом древе грибов соответствует модели неофункционализации С. Оно.

Анализируя данные, представленные на рис. 1, также можно отметить иные, таксон-специфические события накопления атипичных аминокислотных замен. Например, в циклине В3 у беспозвоночных ускорение темпов фиксации атипичных замен выявлено на этапе возникновения двукрылых и перепончатокрылых насекомых, у позвоночных – на этапе возникновения млекопитающих. Следует отметить, что циклины В3 в большей степени важны для контроля мейоза, чем митоза [45–48]. У мышей и нематод, мутантных по циклину В3, наблюдаются нарушения в дифференцировке половых клеток [45–48]. Следовательно, можно предположить, что на этапах возникновения двукрылых и перепончатокрылых насекомых, а также при возникновении млекопитающих происходят существенные изменения в контроле мейоза циклинами В3. Также интересно отметить, что атипичные замены аминокислот в циклине В характерны для эволюции на этапе образования кишечнополостных.

На втором этапе мы проанализировали статистическую взаимосвязь между изменением последовательности циклинов и изменением в ходе эволюции важного фенотипического признака организмов – сложности (см. Материалы и методики). Изменения циклинов измерялись по изменению физико-химических свойств аминокислот в белке в течение эволюции от белка общего предка всех циклинов В до современного белка. Для преобразования филограммы циклинов В (см. рис. 1) в хронограмму использовались 19 палеонтологических границ варьирования датировок дивергенций родов [49–53] (табл. 1).

Для ответа на вопрос, существует ли связь между изменением сложности организмов и эволюцией циклинов, нами был проведен анализ взаимосвязи между эволюционными изменениями свойств аминокислот в трех филетических группах циклинов В (циклины В3 животных, циклины грибов и циклины В1-2 животных) и изменением сложности организмов. Статистический анализ проводился группой методов GLS.

В ходе статистического анализа было выявлено, что с изменением сложности организмов статистически высокозначимо ($p < 0,01$) коррелировали

изменения ряда физико-химических свойств аминокислот *только* в группе циклинов В3 (табл. 2).

Т а б л и ц а 1

Границы варьирования датировок дивергенций родов

Дивергенция	Время дивергенции, млн лет назад		Дивергенция	Время дивергенции, млн лет назад	
	max	min		max	min
Apis – Drosophila	307,2	238,5	Oryzias – Fugu	150,9	96,9
Drosophila – Culex	295,4	238,5	Danio – Fugu	165,2	149,85
Homo – Macaca	33,9	23,0	Danio – Gallus	421,75	416
Mus – Rattus	12,3	11	Xenopus – Gallus	350,1	330,4
Homo – Mus	100,5	61,7	Drosophila – Homo	649	531,5
Homo – Bos	113	71,2	Hydra – Drosophila	712	615
Bos – Canis	71,2	62,2	Fusarium – Magnaporthe	500	400
Ornithorhynchus – Homo	191,1	162,5	Fusarium – Aspergillus	600	400
Taeniopygia – Gallus	86,5	66	Saccharomyces – Candida	350	150
Gallus – Homo	330,4	312,3			

Т а б л и ц а 2

Физико-химические свойства аминокислот циклинов В3, изменение которых коррелирует со сложностью организмов ($p < 0,01$)

Название физико-химического свойства	Группа физико-химического свойства по данным [4]
Нормализованные параметры гибкости В-величин для каждого остатка, окруженного двумя жесткими соседями [54]	Гидрофильность и формирование поверхностных участков белка
Нормализованные параметры гибкости В-величин для каждого остатка, окруженного одним жестким соседом [54]	
Частота 4-го остатка в повороте [55]	Формирование неструктурированных участков белка

Высокозначимая корреляция между изменением сложности организмов и изменением физико-химических свойств не наблюдалась в филетических группах циклинов В1-2 животных и циклинов грибов. В семействе циклинов В3 скоррелировано со сложностью организмов в ходе эволюции изменялись свойства аминокислот, относящиеся к двум группам (см. табл. 1) – группам физико-химических свойств, связанных с гидрофильностью и формированием поверхностных неструктурированных участков белков (петлевые структуры, переходные участки между петлями и β -слоями и α -спиралями).

Иначе говоря, эволюционные изменения свойств аминокислот у циклинов В3, скоррелированные с изменением сложности организмов, указывают на то, что связанная с изменением сложности организмов эволюция этих циклинов происходила преимущественно по неструктурированным, поверхностным участкам белков. В то же время известно, что циклины осуществляют свою функцию опосредованно, за счет белок-белковых взаимо-

действий [8]. Таким образом, по-видимому, в процессе эволюции сложности организмов происходили скоррелированные изменения в числе и природе белок-белковых взаимодействий циклинов В3, играющих существенную роль в контроле мейоза.

Почему с изменением сложности организмов высокозначимо коррелировали изменения физико-химических свойств аминокислот только в группе циклинов В3? У большинства животных, за исключением наиболее простых, известно четыре семейства основных циклинов – А, В, D и Е, контролирующих различные фазы клеточного цикла [8]. Циклин D контролирует фазу G1, сильно варьирующую по длительности в разных клеточных типах; циклин Е контролирует конец фазы G1 и также сильно варьирующий по длительности переход G1/S; циклин А контролирует выход из S-фазы и наиболее консервативную часть клеточного цикла – переход G2/M; циклин В1-2 контролирует только самую консервативную часть клеточного цикла – переход G2/M [8]. Возможно, именно поэтому молекулярная эволюция циклинов В1-2 животных не согласуется с эволюцией сложности животных.

У грибов из четырех семейств основных циклинов, характерных для животных, в геномах представлены только циклины типа В. У *S. pombe* есть только одно семейство этих основных циклинов – *cdc13*. Показано, что *cdc13* способен в одиночку обеспечивать прохождение клетки через все фазы клеточного цикла [56]. У других грибов, например *S. cerevisiae*, есть несколько циклинов семейства В: *clb1-2*, *clb3-4*, *clb5-6*, которые контролируют различные фазы клеточного цикла (*clb5-6* – S фазу, *clb1-2*, *clb3-4* – G2 и M-фазы) [57]. Такое различие по числу представленных в геномах грибов циклинов В, а также отсутствие корреляции между изменением сложности и эволюцией циклинов могут быть связаны с мицелиальной организацией тела грибов, когда процесс эволюционного усложнения организации сводится в первую очередь к изменениям в биохимических путях, и только частично – к изменению формы гифа и нарушению синхронности деления ядер в гифе [58].

Заключение

Создана компьютерная система анализа режимов эволюции генов и белков. В основе системы лежат Марковское моделирование эволюции генов и белков и анализ изменения всех известных физико-химических свойств аминокислот. Уникальной особенностью системы является возможность напрямую соотносить вычисленные особенности молекулярной эволюции с эволюцией заданных фенотипических признаков организмов и, таким образом, легко интерпретировать результат.

С помощью созданной системы проведено исследование режимов молекулярной эволюции циклинов В. Показано, что фиксация атипичных аминокислотных замен идет после дупликации у обоих паралофов В1 и В2 циклинов позвоночных, а также только в одной из дублированных копий в ряду последовательных дупликаций в ходе молекулярной эволюции циклинов В

грибов. В результате корреляционного анализа показано, что у животных эволюция подсемейства циклинов В3, в отличие от подсемейств циклинов В1 и В2, была связана с усложнением организма, что может быть объяснено важной ролью циклинов В3 в контроле мейоза, а следовательно, и в контроле репродуктивной изоляции видов.

Литература

1. Drummond D.A., Bloom J.D., Adami C. et al. Why highly expressed proteins evolve slowly // Proc. Natl. Acad. Sci. U.S.A. 2005. Vol. 102, № 40. P. 14338–14343.
2. Drummond D.A., Wilke C.O. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution // Cell. 2008. Vol. 134, № 2. P. 341–352.
3. Gout J.F., Kahn D., Duret L. Paramecium Post-Genomics Consortium. The relationship among gene expression, the evolution of gene dosage, and the rate of protein evolution // PLoS Genet. 2010. Vol. 6, № 5. P. e1000944.
4. Kawashima S., Pokarowski P., Pokarowska M. et al. AAindex: amino acid index database, progress report 2008 // Nucleic Acids Res. 2008. Vol. 36, Database issue. P. D202–D205.
5. Zhang J. Rates of conservative and radical nonsynonymous nucleotide substitutions in mammalian nuclear genes // J. Mol. Evol. 2000. Vol. 50, № 1. P. 56–68.
6. Smith Nick G.C. Are radical and conservative substitution rates useful statistics in molecular evolution? // J. Mol. Evol. 2003. Vol. 57, № 4. P. 467–478.
7. Pupko T., Sharan R., Hasegawa M. et al. Detecting excess radical replacements in phylogenetic trees // Gene. 2003. Vol. 319, P. 127–135.
8. Cell Cycle Regulation / ed. by P. Kaldis. Berlin : Springer-Verlag, 2006. 374 p.
9. Nieduszynski C.A., Murray J., Carrington M. Whole-genome analysis of animal A- and B-type cyclins // Genome Biol. 2002. Vol. 3, № 12. P. RESEARCH0070.
10. Analysis of Phylogenetics and Evolution with R / ed. by E. Paradis. N.Y. : Springer Science+Business Media, 2006. P. 133–183.
11. Gunbin K.V., Afonnikov D.A., Kolchanov N.A. Molecular evolution of the hyperthermophilic archaea of the Pyrococcus genus: analysis of adaptation to different environmental conditions // BMC Genomics. 2009. Vol. 10. P. 639.
12. Kanehisa M., Goto S., Furumichi M. et al. KEGG for representation and analysis of molecular networks involving diseases and drugs // Nucleic Acids Res. 2010. Vol. 38, Database issue. P. D355–D360.
13. Maddison D.R., Schulz K.-S. (eds.) 2007. The Tree of Life Web Project. URL: <http://tolweb.org>
14. Burki F., Shalchian-Tabrizi K., Minge M. et al. Phylogenomics reshuffles the eukaryotic supergroups // PLoS One. 2007. Vol. 2, № 8. P. e790.
15. Yoon H.S., Grant J., Tekle Y.I. et al. Broadly sampled multigene trees of eukaryotes // BMC Evol. Biol. 2008. Vol. 8. P. 14.
16. Hampl V., Hug L., Leigh J.W. et al. Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic «supergroups» // Proc. Natl. Acad. Sci. U.S.A. 2009. Vol. 106, № 10. P. 3859–3864.
17. Minge M.A., Silberman J.D., Orr R.J. et al. Evolutionary position of breviate amoebae and the primary eukaryote divergence // Proc. Biol. Sci. 2009. Vol. 276, № 1657. P. 597–604.
18. Fitzpatrick D.A., Logue M.E., Stajich J.E., Butler G. A fungal phylogeny based on 42 complete genomes derived from supertree and combined gene analysis // BMC Evol. Biol. 2006. Vol. 6. P. 99.
19. Wang H., Xu Z., Gao L., Hao B. A fungal phylogeny based on 82 complete genomes using the composition vector method // BMC Evol. Biol. 2009. Vol. 9. P. 195.

20. *Dopazo H., Dopazo J.* Genome-scale evidence of the nematode-arthropod clade // *Genome Biol.* 2005. Vol. 6, № 5. P. R41.
21. *Irimia M., Maeso I., Penny D. et al.* Rare coding sequence changes are consistent with Ecdysozoa, not Coelomata // *Mol. Biol. Evol.* 2007. Vol. 24, № 8. P. 1604–1607.
22. *Philippe H., Brinkmann H., Martinez P. et al.* Acoel flatworms are not platyhelminthes: evidence from phylogenomics // *PLoS One.* 2007. Vol. 2, № 1. P. e717.
23. *Lartillot N., Philippe H.* Improvement of molecular phylogenetic inference and the phylogeny of Bilateria // *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 2008. Vol. 363, № 1496. P. 1463–1472.
24. *Marletaz F., Le Parco Y.* Careful with understudied phyla: the case of chaetognath // *BMC Evol. Biol.* 2008. Vol. 8. P. 251.
25. *Helmkamp M., Bruchhaus I., Hausdorf B.* Phylogenomic analyses of lophophorates (brachiopods, phoronids and bryozoans) confirm the Lophotrochozoa concept // *Proc. Biol. Sci.* 2008. Vol. 275, № 1645. P. 1927–1933.
26. *Roy S.W., Irimia M.* Rare genomic characters do not support Coelomata: intron loss/gain // *Mol. Biol. Evol.* 2008. Vol. 25, № 4. P. 620–623.
27. *Podsiadlowski L., Braband A., Struck T.H. et al.* Phylogeny and mitochondrial gene order variation in Lophotrochozoa in the light of new mitogenomic data from Nemertea // *BMC Genomics.* 2009. Vol. 10. P. 364.
28. *Holland L.Z., Albalat R., Azumi K. et al.* The amphioxus genome illuminates vertebrate origins and cephalochordate biology // *Genome Res.* 2008. Vol. 18, № 7. P. 1100–1111.
29. *Swalla B.J., Smith A.B.* Deciphering deuterostome phylogeny: molecular, morphological and palaeontological perspectives // *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 2008. Vol. 363, № 1496. P. 1557–1568.
30. *Singh T.R., Tsagkogeorga G., Delsuc F. et al.* Tunicate mitogenomics and phylogenetics: peculiarities of the *Herdmania momus* mitochondrial genome and support for the new chordate phylogeny // *BMC Genomics.* 2009. Vol. 10. P. 534.
31. *Hallstrom B.M., Kullberg M., Nilsson M.A., Janke A.* Phylogenomic data analyses provide evidence that Xenarthra and Afrotheria are sister groups // *Mol. Biol. Evol.* 2007. Vol. 24, № 9. P. 2059–2068.
32. *Kitazoe Y., Kishino H., Waddell P.J. et al.* Robust time estimation reconciles views of the antiquity of placental mammals // *PLoS One.* 2007. Vol. 2, № 4. P. e384.
33. *Kjer K.M., Honeycutt R.L.* Site specific rates of mitochondrial genomes and the phylogeny of eutheria // *BMC Evol. Biol.* 2007. Vol. 7. P. 8.
34. *Murphy W.J., Pringle T.H., Crider T.A. et al.* Using genomic data to unravel the root of the placental mammal phylogeny // *Genome Res.* 2007. Vol. 17, № 4. P. 413–421.
35. *Hou Z.C., Romero R., Wildman D.E.* Phylogeny of the Ferungulata (Mammalia: Laurasiatheria) as determined from phylogenomic data // *Mol. Phylogenet. Evol.* 2009. Vol. 52, № 3. P. 660–664.
36. *Krauss V., Thummler C., Georgi F. et al.* Near intron positions are reliable phylogenetic markers: an application to holometabolous insects // *Mol. Biol. Evol.* 2008. Vol. 25, № 5. P. 821–830.
37. *Kolaczkowski B., Thornton J.W.* A mixed branch length model of heterotachy improves phylogenetic accuracy // *Mol. Biol. Evol.* 2008. Vol. 25, № 6. P. 1054–1066.
38. *Archambault V., Buchler N.E., Wilmes G.M. et al.* Two-faced cyclins with eyes on the targets // *Cell Cycle.* 2005. Vol. 4, № 1. P. 125–130.
39. *Vogel C., Chothia C.* Protein Family Expansions and Biological Complexity // *PLoS Comput. Biol.* 2006. Vol. 2, № 5. P. e48.
40. *Wilson D., Pethica R., Zhou Y. et al.* SUPERFAMILY – Comparative Genomics, Datamining and Sophisticated Visualisation // *Nucleic Acids Res.* 2009. Vol. 37, Database issue. P. D380–D386.

41. *Hughes A.L.* The evolution of functionally novel proteins after gene duplication // *Proc. R. Soc. Lond. Ser. B Biol. Sci.* 1994. Vol. 256, № 1346. P. 119–124.
42. *Brandeis M., Rosewell I., Carrington M. et al.* Cyclin B2-null mice develop normally and are fertile whereas cyclin B1-null mice die in utero // *Proc. Natl. Acad. Sci. U.S.A.* 1998. Vol. 95, № 8. P. 4344–4349.
43. *Оно С.* Генетические механизмы прогрессивной эволюции. М. : Мир, 1973. 227 с.
44. *Teichmann S.A., Babu M.M.* Gene regulatory network growth by duplication // *Nat. Genet.* 2004. Vol. 36, № 5. P. 492–496.
45. *van der Voet M., Lorson M.A., Srinivasan D.G. et al.* *C. elegans* mitotic cyclins have distinct as well as overlapping functions in chromosome segregation // *Cell Cycle.* 2009. Vol. 8, № 4. P. 4091–4102.
46. *Miles D.C., van den Bergen J.A., Sinclair A.H., Western P.S.* Regulation of the female mouse germ cell cycle during entry into meiosis // *Cell Cycle.* 2010. Vol. 9, № 2. P. 408–418.
47. *Nguyen T.B., Manova K., Capodiceci P. et al.* Characterization and expression of mammalian cyclin b3, a prepachytene meiotic cyclin // *J. Biol. Chem.* 2002. Vol. 277, № 44. P. 41960–41969.
48. *Refik-Rogers J., Manova K., Koff A.* Misexpression of cyclin B3 leads to aberrant spermatogenesis // *Cell Cycle.* 2006. Vol. 5, № 17. P. 1966–1973.
49. *Benton M.J., Donoghue P.C.* Paleontological evidence to date the tree of life // *Mol. Biol. Evol.* 2007. Vol. 24, № 1. P. 26–53.
50. *Peterson K.J., Cotton J.A., Gehling J.G., Pisani D.* The Ediacaran emergence of bilaterians: congruence between the genetic and the geological fossil records // *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 2008. Vol. 363, № 1496. P. 1435–1443.
51. *Budd G.E.* The earliest fossil record of the animals and its significance // *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 2008. Vol. 363, № 1496. P. 1425–1434.
52. *Taylor J.W., Berbee M.L.* Dating divergences in the Fungal Tree of Life: review and new analyses // *Mycologia.* 2006. Vol. 98, № 6. P. 838–849.
53. *Berbee M.L., Taylor J.W.* Dating the molecular clock in fungi – how close are we? // *Fungal Biol. Rev.* 2010. Vol. 24, № 1–2. P. 1–16.
54. *Vihinen M., Torkkila E., Riihonen P.* Accuracy of protein flexibility predictions // *Proteins.* 1994. Vol. 19, № 2. P. 141–149.
55. *Chou P.Y., Fasman G.D.* Prediction of the secondary structure of proteins from their amino acid sequence // *Adv. Enzymol.* 1978. Vol. 47. P. 45–148.
56. *Fisher D.L., Nurse P.* A single fission yeast mitotic cyclin B p34cdc2 kinase promotes both S-phase and mitosis in the absence of G1 cyclins // *EMBO J.* 1996. Vol. 15, № 4. P. 850–860.
57. *Bloom J., Cross F.R.* Multiple levels of cyclin specificity in cell-cycle control // *Nat. Rev. Mol. Cell Biol.* 2007. Vol. 8, № 2. P. 149–160.
58. *Мюллер Э., Леффлер В.* Микология. М. : Мир, 1995. 343 с.

Поступила в редакцию 03.12.2010 г.

Konstantin V. Gunbin¹, Mikhail A. Genaev¹, Igor I. Turnaev¹, Dmitry A. Afonnikov^{1,2}

¹Institute of Cytology and Genetics of Siberian Branch
of Russian Academy of Sciences, Novosibirsk, Russia

²Novosibirsk State University, Novosibirsk, Russia

COMPUTER SYSTEM FOR ANALYSIS OF MOLECULAR EVOLUTION MODES OF GENES AND PROTEINS: A CASE STUDY OF THE CYCLIN B MOLECULAR EVOLUTION

The evolution of protein-coding genes is characterized by two types of single-nucleotide substitutions: non-synonymous, altering protein structure, and synonymous, that does not change the protein. Natural selection of proteins is manifested through various non-synonymous fixations. Protein function, structure and folding are all determined by the combination of physicochemical properties of amino acids. To analyze the protein evolution mode two main approaches were previously introduced: (1) based on the calculation of the radical (large changes in certain physicochemical property) to conservative (small changes in certain physicochemical property) codon substitutions rate ratio (K_R/K_C), and (2) based on analysis of variance of rate changes of various physicochemical amino acid properties (V_{PC}) in the course of protein evolution. The essential drawback of these approaches is the need to know in advance “positive change” of certain amino acid properties.

Our novel computer system for protein evolution mode analysis is also based on the K_R/K_C and V_{PC} , but has two crucial differences which allow us to overcome the drawback: (1) we take into account all the known properties of amino acids, and (2) we calculate the statistical relation of properties changes with certain adaptive phenotypical features of organisms. The opportunity to match evolutionary change of all amino acid properties with positively selected phenotypical features of organisms permits direct attribution of certain protein changes to selection events. On this basis, a computer system for analyzing of molecular evolution mode of genes and proteins was developed (<http://pixie.bionet.nsc.ru/samem/>). Our computer system consists of two main pipelines, analyzing the evolutionary modes of genes (I) and proteins (II), and two supplementary pipelines, collecting a sample of genes and proteins (III) and producing their primary analysis (IV). An essential feature of the gene analysis pipeline is a possibility to study gene evolution modes using different K_R/K_C calculation methods. Pipeline for protein analysis implements a novel algorithm for detection of atypical amino acids fixations on phylogenetic tree branches. This algorithm is based on Markov simulation of protein sequence evolution, and uses the permutation test for comparison of the simulated protein molecular evolution with real one.

To test the capabilities of our computer system and performance of our novel algorithm for detection of atypical amino acids fixations on tree branches the molecular evolution of cyclin B protein family was analyzed. It was shown that the fixation of atypical amino acid substitutions is frequent to both cyclin paralogs B1 and B2 after cyclin B duplication in the vertebrate evolution. In contrast, only one of the duplicated cyclin B copies possessed of atypical amino acid fixations in a series of successive cyclin B duplications in the fungal evolution. Moreover, it was shown that the evolution of cyclins B3 subfamily, in contrast with cyclin B1 and B2 subfamilies, was strictly associated with increasing of animal's complexity. This association can be explained by the important role of cyclin B3 in the meiosis and, hence, in the mechanisms of reproductive isolation.

Key words: molecular evolution; conservative and radical amino acid substitutions; client-server system; statistical analysis; cyclin B.

Received December 3, 2010