

Отзыв официального оппонента

на диссертационную работу

Базыкина Георгия Александровича

«Положительный и эпистатический отбор в эволюции аминокислотных последовательностей»

представленную на соискание ученой степени доктора биологических наук по специальности «03.01.09 – математическая биология, биоинформатика»

Диссертационная работа Г.А.Базыкина представляет фундаментальное исследование по эволюционной биологии – исследование отбора, действующего на аминокислотные сайты белок-кодирующих генов, и изменения характеристик этого отбора в ходе эволюции. Основной теоретический объект исследования - поверхность приспособленности (адаптивный ландшафт) — функция, ставящая в соответствие генотипам приспособленность. Автор представляет новое понятие однопозиционного адаптивного ландшафта (ОПАЛ) — набора значений приспособленности аминокислот в одном аминокислотном сайте, показывает применение данного подхода для решения задач эволюционной биологии. Концепция адаптивного ландшафта развивается автором далее в диссертации, рассматриваются изменения ОПАЛ в ходе эволюции, изменение скорости и направления замен, их взаимозависимости. Было известно, что действие отбора в аминокислотном сайте часто зависит от аминокислот, присутствующих в других сайтах того же или другого белка. Однако систематическое исследование этого явления, требовало развития новых биоинформационных и статистических методов, приведенных в представленной Г.А.Базыкиным работе. Все исследования выполнены на больших выборках полногеномных данных для большого набора видов - от вирусов и одноклеточных - до человека.

Продолжающееся быстрое развитие высокопроизводительных постгеномных технологий, методов экспериментального определения

аминокислотных последовательностей дает возможности расширения наших знаний об эволюции, представляя широкую область применения разработанной теории.

Актуальность исследования. Систематическое исследование геномов и протеомов требует как разработки новых методов обработки геномных данных, так и развития теоретических концепций, которые позволили бы получить новые знания об эволюции на основе накопленных данных. Проект “Геном человека” позволил систематизировать представления о структуре генома и дал начало новой, постгеномной эпохе, развитию современных омиксных технологий. Однако остаются нерешенными многие задачи эволюционной геномики, вопросы изменения генотипа в ходе эволюции, которые требуют новых теоретических подходов, в том числе оценки изменения «ландшафта» - частот нуклеотидов и аминокислот в ходе адаптации. Актуальным является анализ последовательностей полных геномов многих видов и индивидов внутри одного вида, для понимания механизмов отбор, действующего на аминокислотные последовательности белок-кодирующих генов. Как отмечает Г.А.Базыкин, ранние исследования показали, что на уровне нуклеотидных последовательностей преобладает отрицательный отбор, и многие аминокислотные позиции эволюционируют эффективно нейтрально. При этом было известно, что часть замен между видами происходила под действием положительного отбора; однако оценки доли замен, происходивших под действием такого отбора между разными видами существенно различались. Это, в свою очередь, требовало разработки новых методов анализа отбора. Анализ доступных в настоящее время полногеномных данных многих видов позволяет изучать редкие явления или сочетания явлений, которые поясняют различные аспекты межгенных взаимодействий, что имеет, в том числе, прикладное значение для ряда задач медицинской генетики (разделы диссертации по эволюции вируса гриппа).

Таким образом актуальность данной диссертационной работы не вызывает сомнения.

Структура и основное содержание диссертации. Диссертация построена традиционным образом и состоит из Введения, четырех Глав: «Обзор литературы», «Адаптивный ландшафт одиночных аминокислотных сайтов», «Эпистатический адаптивный ландшафт пар аминокислотных сайтов», «Эпистатический отбор на концах кодирующей части гена», заключения и списка используемой литературы содержащего 429 источников. Диссертационная работа изложена на 199 страницах машинописного текста, включая 29 рисунков и 20 таблиц.

Во Введении автор описывает актуальность проблемы, цель и задачи исследования, а также обсуждает научную новизну и практическую ценность развития методов. Данный раздел содержит все необходимые подразделы, включая актуальность темы, ее разработанность, значимость, постановку задач исследования, новизну и положения, выносимые на защиту, апробацию результатов и формальные характеристики работы.

Основное содержание диссертации представлено четырьмя главами. В главе «Обзор литературы» автор определяет понятие ОПАЛ и приводит характеристики однопозиционного адаптивных ландшафтов (ОПАЛов) в зависимости от их изменчивости в заданный момент времени и в ходе эволюции. На моделях инвариантных и изменчивых эволюционных ландшафтов Г.А. Базыкин рассматривает вероятности изменения ОПАЛов, действие на них положительного и отрицательного отбора.

Проиллюстрированы возможные механизмы формирования гомоплазий (многократных возникновений одного и того же признака — в данной работе - аминокислоты) и скоррелированной генетической изменчивости. В завершение Г.А.Базыкин определяет цель и задачи исследования, связанные с оценками эволюционной изменчивости ОПАЛов.

Автор приводит убедительный список публикаций - 40 статей, все в журналах, рекомендованных ВАК. Однако во Введении отмечает «Материал 23 из этих статей имеет прямое отношение к тексту диссертации». Заметим,

что было бы достаточным указать только эти 23 работы, отметив, что методы были представлены в большем числе публикаций.

Последовательность и логика представленных результатов указаны в каждом разделе, что облегчает восприятие материала диссертации. Однако язык изложения достаточно лаконичен, и Глава «Обзор литературы» могла бы быть расширена в объеме (без увеличения числа цитирований), до 40-50 страниц.

Глава «Адаптивный ландшафт одиночных аминокислотных сайтов» детально описывает концепцию ОПАЛов. Консервативные сегменты кодирующей и не кодирующей ДНК эволюционируют медленно из-за действия отрицательного отбора, удаляющего новые мутации. Однако какой отбор действует на те замены, которые все же происходят в таких сегментах, остается неясным. Г.А.Базыкин показывает, что в генах, кодирующих белки, у *Drosophila* доля аллельных замещений, вызывавшихся положительным отбором, а также сила этого отбора максимальны в консервативных сегментах генов. Адаптивный ландшафт локуса, то есть вектор приспособленности аллелей в нем, может изменяться в результате аллельных замещений в других локусах в присутствии эпистатических взаимодействий между локусами. В паре дивергирующих гомологичных белков вероятность того, что аминокислотная замена в одном из них сделает его более похожим на другой, исходно высока, однако снижается со временем, что свидетельствует о дивергенции адаптивных ландшафтов гомологичных сайтов.

Глава «Эпистатический адаптивный ландшафт пар аминокислотных сайтов» расширяет исследования изменений на пары сайтов в геноме, вообще говоря не обязательно относящихся к одному и тому же белку. Ранее Г.А.Базыкин показал, что положительный эпистаз проявляется как неравномерность распределения замен по филогении: пара замен, находящихся в положительном эпистазе друг с другом, чаще происходит быстро одна за другой, чем ожидается в случае их независимости.

В следующей Главе - «Эпистатический отбор на концах кодирующей части гена» автор использует ту же логику для исследования взаимодействий между различными эволюционными событиями. Эта Глава диссертационной работы посвящена эволюции аминокислотных сайтов под эпистатическим отбором в специальном случае — на краях кодирующей части генов.

Г.А.Базыкин сравнил два гаплоидных генотипа одного индивида асцидии *Ciona savignyi* и определил кодоны, в которых эти генотипы отличаются двумя несинонимическими заменами.

Альтернативные стартовые кодоны AUG в пределах одного транскрипта могут вносить вклад в разнообразие протеома, однако их функциональная значимость оставалась спорной. Автор представил доказательства того, что альтернативные старт-кодоны находятся под действием отрицательного отбора у позвоночных, насекомых и дрожжей.

Степень обоснованности и достоверность научных положений и выводов. Работа выполнена на высоком методическом и техническом уровне. Используются актуальные версии баз данных и репозиториев, отражающих все имеющиеся на данный момент знания о белках и генах. Выводы работы полностью обоснованы и логически вытекают из полученных результатов.

Материалы диссертации опубликованы в более чем достаточном числе статей в высокорейтинговых рецензируемых международных научных журналах (23 статьи, все по списку ВАК), представлены на международных и отечественных научных конференциях, что не вызывает сомнений в проработанности и достоверности научных положений..

Новизна исследований и полученных результатов. В данной работе впервые предложена концепция однопозиционных адаптивных ландшафтов (ОПАЛ) — векторов приспособленности различных аминокислот в определенном сайте белка. Предложен подход к оценке доли аминокислотных замен, происходивших при поддержке положительного естественного отбора, основанный на сопоставлении параметров

внутривидовой изменчивости в функционально важных и селективно нейтральных сайтах. Выявлены события реассортации, происходившие между восемью сегментами вируса гриппа H3N2 после его выявления 1968 году.

Автором показано, что поверхностные белки гемагглютинин (HA) и нейраминидаза (NA) вируса гриппа А человека эволюционируют под давлением отбора, направленного на избежание ответа адаптивной иммунной системы и действия противовирусных препаратов. Показано, что альтернативные старт-кодоны находятся под действием отрицательного отбора у позвоночных, насекомых и дрожжей.

Описана эволюция длины кодирующей последовательности прокариотических генов в результате изменения положения стоп-кодонов: случаи добавления участков 3'-нетранслируемой последовательности к генам в результате мутаций существующих стоп-кодонов и случаи исключения С-концевых кодирующих сегментов в результате нонсенс-мутаций выше стоп-кодона.

Разработан новый филогенетический метод для определения следов генетических взаимодействий между мутациями в разных генах, то есть межгенного эпистаза, учитывающего возможные реассортации. С использованием этого метода показано, что поверхностные белки гриппа эволюционируют скоординированным образом, так что замены в HA влияют на замены в NA во многих сайтах, и наоборот.

Теоретическая и практическая значимость полученных результатов. Исследование имеет фундаментальный характер, определяя новые подходы и метод сравнительной эволюционной биологии. В плане приложений рассматривается структура и динамика адаптивных ландшафтов в природных популяциях. Разработанный тест на положительный отбор применим в широком классе случаев, когда имеются данные по межвидовой дивергенции и по внутрипопуляционному полиморфизму позволяет оценить долю замен, которым способствовал положительный отбор, среди всех

аминокислотных замен, отдельно для различных этапов эволюционной истории вида. Оценка этой доли для модельных видов — человека и дрозофилы — проливает свет на историю адаптивной эволюции. Показанное автором увеличение доли адаптивных замен с консервативностью аминокислотного сайта меняет существующие представления, согласно которым адаптивная эволюция имеет наибольшее значение в быстро эволюционирующих сайтах.

Описание скоординированной эволюции поверхностных белков вируса гриппа является первым систематическим описанием положительного межгенного эпистаза по реконструированным заменам, происходившим в ходе эволюции; однако разработанный с этой целью метод применим для поиска положительного эпистаза и в других системах с полным или частичным сцеплением между анализируемыми локусами.

С точки зрения практической значимости понимание структуры отбора на аминокислотные сайты патогенов и роли взаимодействий между сайтами в определении этого отбора необходимо для оптимизации профилактики и лечения заболеваний, вызываемых этими патогенами, в том числе оптимизации выбора лекарственных средств и штаммов для вакцинации. Описанные взаимодействия с участием мутации устойчивости к озельтамивиру (лекарственному препарату Тамифлю) у вируса гриппа проливают свет на пути возникновения и распространения лекарственной устойчивости. Описанная связь между реассортациями и последующим антигенным дрейфом важна для моделирования последующего возникновения новых штаммов, в т.ч. для оценки вероятности распространения патогенного штамма вируса гриппа H5N1.

Замечания и рекомендации.

Среди достоинств представленной работы необходимо отметить актуальность проблемы, новизну подхода, тщательную методологическую

проработку задачи. Однако представляется целесообразным высказать ряд пожеланий и замечаний к работе.

Прежде всего работа относится к биоинформатике и предполагает разработку компьютерных программ. Никакого упоминания о программах, языке программирования, их доступности в тексте нет. Хотя из ссылок на статьи автора ясно, что были разработаны собственные оригинальные компьютерные программы (возможно коллективом авторов с участием Г.А.Базыкина).

Было бы уместно привести часть кода, основные блок-схемы в Приложении. Разработка компьютерных программ, их использовании в биоинформатике стоило привести в разделе «Практическая значимость», тем более соискатель ведет преподавательскую работу, работает в составе научной группы. Ведь результат работы не только теоретический.

При исследовании пар взаимодействующих позиций не упомянуты явно работы по исследованию коморбидности – совместной встречаемости заболеваний, и соответственно совместных мутаций в парах генов, взаимодействующих (либо в одном метаболическом пути, либо опосредованным образом). Расстояние между генами в геноме также имеет значение часто свидетельствуя об общей биологической функции. Тем более расстояние между позициями аминокислот внутри последовательности одного белка с ограничениями из элементов вторичной и пространственной структуры.

Стоило бы упомянуть работы C.R.Sanders, например - Schleich JP, Sanders CR. «The safety dance: biophysics of membrane protein folding and misfolding in a cellular context» Q Rev Biophys. 2015.

Sakakura M, ..., Sanders CR. «Structural basis for the Trembler-J phenotype of Charcot-Marie-Tooth disease» Structure. 2011.

Если в консервативном участке на изменения аминокислот действует отрицательный отбор, чтобы сохранить структуры белка, то что меняется –

активный сайт белка, или время денатурации и агрегации с другими белками. Такие компенсаторные эффекты выходят за рамки представленной работы, но представляется интересным их обсудить в дискуссии в плане дальнейшего развития направления исследований.

Концепция ОПАЛ приведена на схематическом рисунке, при этом аминокислоты отмечены просто номерами. Если сгруппировать аминокислоты вместе по физико-химическим свойствам, например от гидрофобных к гидрофильным, то будут ли видны новые закономерности (ограничения на переход в адаптивном ландшафте - в консервативных сайтах замены в противоположные по свойствам аминокислоты будут невозможны, либо наоборот усилятся)? Было бы уместно привести картины переходов в ОПАЛ по консервативным сайтам, уже рассчитанные для *Drosophila* отдельным рисунком, или в Приложении.

Ряд сокращений даны без пояснения, которое приводится в тексте позже, в частности НА и NA для гемагглютинаина (НА) и нейраминидазы (NA). Обозначения dAUG и sAUG, относящиеся к стоп-кодону AUG в 3'- и 5'-районе гена даны только в автореферате, и не в первом упоминании в тексте.

Выводы не должны содержать сокращений, вновь НА и NA даны без пояснений.

Автореферат, содержание которого соответствует диссертации, не превышает стандарт 32 страницы. В то же время в автореферате дано обширное цитирование литературы из текста основной диссертации. Было бы уместно дать краткий список цитируемой литературе в автореферате, или избежать по возможности цитирование, или приводить названия цитируемых статей и журнал прямо в тексте.

Важно каждое положение и вывод подтверждать ссылками на собственные публикации, тем более они есть. Иначе не всегда ясно, в какой именно работе автора доказано, то или иное утверждение.

Есть незначительные опечатки в тексте диссертации.

«Материал 23 из этих статей имеет прямое отношение к тексту диссертации». Полный список 40 публикаций автора так и не приведен в диссертации, дано 29 статей. Эту фразу следовало бы поправить или дать полный список публикаций автора (в тексте диссертации), включая все методические работы.

В работах принято говорить о роли автора во множественном числе (указывая на работу коллектива), в то же время в диссертации есть фразы:

«В главе 1, ... я описываю текущее состояние...»

«В главе 2 я разрабатываю новые методы анализа...»

«В главе 3 мы используем ту же логику...»

В то же время автореферате таких фраз нет, все корректно.

Порядок цитирования - надо приводить по годам (стр.31) - (напр. Mustonen, Lässig 2007; Povolotskaya, Kondrashov 2010; Breen et al. 2012; Fitch, Markowitz 1970)

Было бы логично процитировать сначала более старые работы, переходя к более новым - 1970, 2007, 2010, 2012.

Формулы не выделены номерами, в частности на стр.36.

На стр. 58 также приведены формулы без номеров (но со ссылками на другие работы)

Стр. 44 - UCSC Genome Bioinformatics Site.

Логичнее назвать этот ресурс UCSC Genome Browser, или Table Browser, если загружались табличные данные

«Генбанка» - правильнее GenBank или Генбанк.

«консервативных сегментах» - явная опечатка – «консервативных».

«в качестве аутгруппа» - в качестве аутгруппы (стр.51), видимо и далее по тексту, должно быть в женском роде – аутгруппа, некорректно с английского.

Отметим, что число опечаток очень невелико для такой сложной работы.

В целом язык изложения очень емкий и краткий, тяжел для восприятия. Так, используются специальные термины, которые требуют дополнительного

пояснения - «генетический контекст» (контекст последовательности или генотипа?). «Синергический (сужающий) эпистаз» - взаимодействия между позициями, но что такой сужающей – смягчающий воздействие?

«Плоский ландшафт» - надо видеть рисунок, чтобы понять, что речь идет о постоянном значении во время эволюции. Стоит давать такие термины в кавычках, или пояснять дополнительно при первой упоминании в тексте.

Стоит приводить цитируемые формулы в работе явно, не ограничиваясь только цитатой – например (см. Kimura 1985, уравн. 3.14). То есть читатель должен найти работу и формулу в ней, затрудняет восприятие.

В работе несколько раз упоминается озельтамивир, однако не дана расшифровка или пояснение, например так: «Противовирусное средство, лекарство от гриппа Озельтамивир (в другой транскрипции — Осельтамавир, торговое название — Тамифлю)». Возможно, не я один сразу не понял, что речь идет об известном препарате Тамифлю.

Сделанные замечания не снижают общей высокой оценки диссертационной работы.

Общее заключение.

Диссертация Базыкина Георгия Александровича представляет собой полноценную завершенную научно-исследовательскую работу, описывающее решение крупной научной проблемы эволюционной биоинформатики, имеющей большое теоретическое и практическое значение. Содержание диссертации полностью отражено в публикациях автора в журналах, рекомендованных ВАК (23 публикации), а также в других рецензируемых изданиях, и представлено на международных и российских научных конференциях. Основные результаты и их интерпретация представлены в автореферате, который адекватно отражает содержание диссертации. Поставленные в диссертации задачи решены полностью.

Таким образом, диссертационная работа Базыкина Георгия Александровича «Положительный и эпистатический отбор в эволюции аминокислотных последовательностей» является законченной научно-

квалификационной работой и отвечает требованиям п.9 Положения "О порядке присуждения ученых степеней" N 842 от 24 сентября 2013 г., с внесенными изменениями от 21 апреля 2016 г. № 335, предъявляемым к диссертациям, а ее автор заслуживает присуждения искомой степени доктора биологических наук по специальности 03.01.09 – математическая биология, биоинформатика.



Орлов Юрий Львович

Доктор биологических наук,

Профессор РАН,

Старший научный сотрудник

Лаборатория эволюционной биоинформатики и теоретической генетики
Федеральное государственное бюджетное научное учреждение
«Федеральный исследовательский центр Институт цитологии и генетики
Сибирского отделения Российской академии наук» (ИЦиГ СО РАН)

630090, Новосибирск, Россия, пр.ак.Лаврентьева,10

Тел./факс: +7(383)3331278

Эл.почта: orlov@bionet.nsc.ru

01.06.2018 г.

Подпись Ю.Л. Орлова заверяю

к.б.н. Орлова Галина Владимировна, к.б.н., ученый секретарь Федерального государственного бюджетного научного учреждения «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук»

