

На правах рукописи

Клинк Галина Викторовна

**Расположение аминокислотных замен на эволюционном дереве как
показатель изменчивости однопозиционного адаптивного ландшафта**

03.01.09 – математическая биология, биоинформатика

Автореферат на соискание ученой степени кандидата биологических наук

Москва – 2020

Работа выполнена в Секторе молекулярной эволюции Федерального государственного бюджетного учреждения науки Института проблем передачи информации им. А.А. Харкевича Российской академии наук (ИППИ РАН)

Научный руководитель: Базыкин Георгий Александрович, доктор биологических наук, Федеральное государственное бюджетное учреждение науки Институт проблем передачи информации им. А.А. Харкевича Российской академии наук

Официальные оппоненты:

Трифонов Владимир Александрович, доктор биологических наук, Институт молекулярной и клеточной биологии Сибирского отделения Российской академии наук, Новосибирск, заведующий лабораторией сравнительной геномики.

Поволоцкая Инна Сергеевна, PhD (признаваемый в РФ) по специальности Биомедицина, Обособленное структурное подразделение "Научно-исследовательский клинический институт педиатрии имени академика Ю.Е. Вельтищева" ФГАОУ ВО РНИМУ им.Н.И.Пирогова, заведующая лабораторией клинической геномики и биоинформатики

Ведущая организация:

Федеральное бюджетное учреждение науки «Центральный НИИ эпидемиологии и Роспотребнадзора»

Защита состоится « » ... 20 г. в ... на заседании диссертационного совета по защите докторских и кандидатских диссертаций Д 002.077.04 (утвержден Приказом Минобрнауки России от 16 декабря 2013 года №978/нк) при Федеральном государственном бюджетном учреждении науки Институте проблем передачи информации им. А.А. Харкевича Российской академии наук (ИППИ РАН) по адресу: 127051, г. Москва, Большой Каретный переулок, д.19, стр. 1., факс: +7 (495) 650-05-79, e-mail: gir@iitp.ru

С диссертацией можно ознакомиться в библиотеке Федерального государственного бюджетного учреждения науки Института проблем передачи информации им. А.А. Харкевича Российской академии наук (ИППИ РАН) и на сайте института:

<http://iitp.ru/ru/dissertation/1391.htm>

Автореферат разослан «__» _____ 202__ г.

Ученый секретарь диссертационного совета

д.б.н., профессор Г.И. Рожкова

Общая характеристика работы

Актуальность темы исследования

Эволюция последовательностей ДНК всех организмов происходит под давлением естественного отбора. Его действие заключается в том, что более приспособленные к условиям среды особи дают больше потомства. Число потомков, оставляемых особью, – ее приспособленность – может быть измерена в ходе трудоёмких экспериментов. Отбор действует на фенотип, но эволюция происходит за счёт изменения генотипа. Функция, связывающая генотип и приспособленность его обладателя, называется адаптивным ландшафтом, или ландшафтом приспособленности. Полный адаптивный ландшафт отражает приспособленность организма при всех сочетаниях нуклеотидов во всех позициях генома. Зная эту функцию, можно не только понять законы эволюции в прошлом и настоящем, но и предугадывать её возможные пути в будущем. Понимание устройства адаптивных ландшафтов может быть очень важно для определения эффекта мутаций у человека, изучения динамики эволюции патогенов, оценки состояния популяций, поиска функциональных взаимодействий между сайтами белков. Однако ландшафт приспособленности многомерен: его размерность равна числу позиций в рассматриваемой последовательности. Вместо полного адаптивного ландшафта можно изучать его срезы. При этом часть информации теряется, но снижается число возможных сочетаний вариантов. Однако, даже ландшафт для аминокислотных позиций одного белка длиной в 100 аминокислот имеет размерность 100 и содержит приспособленности для 20^{100} возможных вариантов последовательности. Получение и изучение таких

ландшафтов – непосильная задача для существующих методов и вычислительных мощностей.

Минимальный адаптивный ландшафт, который имеет смысл рассматривать в отдельности, – это однопозиционный адаптивный ландшафт – ОПАЛ. ОПАЛ каждого сайта можно представить в виде вектора длиной 20, где для каждой аминокислоты указано, какой эффект она окажет на приспособленность, если будет стоять в этом сайте. Будем называть этот эффект приспособленностью аминокислоты в данном сайте. ОПАЛ сайта может меняться во времени из-за изменений в геномном контексте: приспособленность одной и той же аминокислоты в сайте может зависеть от того, какие аминокислоты стоят в других позициях белка. Тогда говорят, что аминокислотные позиции эпистатически взаимодействуют. Также ОПАЛ сайта, как и адаптивный ландшафт любой размерности, зависит от факторов среды, в которой функционирует белок [Bazykin 2015].

Насколько сходны ОПАЛы одних и тех же сайтов у разных видов? Как точно можно предсказать ОПАЛ сайта для одного вида живых существ по ОПАЛу для другого вида? Изучение изменчивости ОПАЛов в молекулярной эволюции – ключ к пониманию поведения сложных, многомерных адаптивных ландшафтов.

Для некоторых белков бактерий, вирусов и дрожжей методами насыщающего мутагенеза показано влияние каждой аминокислотной замены в каждом сайте на приспособленность организма [Firnberg и др. 2016, Haddox и др. 2016, Roscoe и др. 2013]. Сравнение полученных на основании таких экспериментов матриц предпочитаемых аминокислот для гомологичных белков в разных видах позволяет сказать, насколько разнятся однопозиционные адаптивные ландшафты их аминокислотных позиций.

Однако изучение изменения приспособленности аминокислот экспериментальными методами возможно не для всех организмов. Кроме того, отбор в условиях лаборатории в некоторых аспектах отличается от отбора в природе [Haddox и др. 2016]. Разработка методов изучения ОПАЛов, основанных на анализе данных секвенирования, необходима для ответа на фундаментальные вопросы об эволюции белков, а также для изучения ландшафтов приспособленности белков человека и его патогенов.

Вирусы – самые быстро эволюционирующие патогены. Большой проблемой для лечения вирусных заболеваний является приобретение этими патогенами устойчивости к противовирусным препаратам [Rhee и др. 2015]. Появление в вирусной популяции отбора на устойчивость к лекарствам меняет ОПАЛы сайтов вирусных белков [Bloom и др. 2010, Flynn и др. 2017]. Таким образом, поиск и изучение аминокислотных сайтов с переменным адаптивным ландшафтом в вирусных белках позволит глубже понять механизмы приобретения вирусами устойчивости к иммунитету и лекарствам.

Данная работа посвящена изучению изменчивости однопозиционных адаптивных ландшафтов филогенетическими методами.

Цели и задачи исследования

Цель исследования – оценить роль изменчивости однопозиционных адаптивных ландшафтов в эволюции белков. Для достижения этой цели поставлены задачи:

1. Оценить изменчивость ОПАЛов митохондриальных белков с помощью анализа расположения аминокислотных замен на филогенетических деревьях.

2. Изучить, как меняется приспособленность референтных, нейтральных и патогенных аллелей митохондриальных белков человека с увеличением филогенетического расстояния от него.

3. Разработать метод поиска аминокислот, приспособленность которых неодинакова у разных видов. С его помощью найти сайты с переменным адаптивным ландшафтом в вирусных белках.

Новизна и практическая значимость работы

Появление одной и той же аминокислоты в конкретном сайте белка может неодинаково отразиться на его функциональности в разных организмах. В таких случаях говорят, что ОПАЛ сайта изменчив. Эволюционная роль этого явления подтверждается с помощью многих методов. Например, показано, что замены на определённую аминокислоту (гомоплазии) распределены на филогенетическом дереве неравномерно. Действительно, частое возникновение определённой аминокислоты в сайте белка в пределах одной клады и отсутствие замен на неё в этом же сайте в другой кладе может говорить о том, что в первой кладе приспособленность этой аминокислоты выше. Для разных типов гомоплазий: реверсий, конвергентных и параллельных замен показана кластеризация на филогении относительно дивергентных замен (то есть, замен на разные аминокислоты) [Goldstein и др. 2015, Naumenko и др. 2012, Povolotskaya и Kondrashov 2010, Rogosin и др. 2008, Zou и Zhang 2015]. В отличие от существующих методик анализа взаиморасположения аминокислотных замен, разработанный нами метод поиска изменения ОПАЛов [Klink и Bazykin 2017] устойчив к кластеризации

гомоплазий в близких видах, не связанной с изменением адаптивного ландшафта.

Ранее было показано, что вероятность наблюдать патогенную для человека аминокислоту как аллель дикого типа в другом биологическом виде либо не зависит от эволюционного расстояния до линии человека [Kondrashov и др. 2002, Soylemez и Kondrashov 2012], либо возрастает с этим расстоянием [Jordan и др. 2015]. Мы впервые показали, что в митохондриальных белках замены на известные патогенные для человека варианты могут кластеризоваться в видах, эволюционно близких к человеку [Klink и др. 2017]. Это наблюдение может помочь меньше ошибаться в интерпретации филогенетической информации при предсказании эффектов мутаций для человека и других организмов.

В данной работе мы не только изучили степень непостоянства однопозиционных адаптивных ландшафтов статистически, но и разработали метод поиска отдельных аминокислот, меняющих приспособленность между видами, основанный на анализе данных секвенирования. С помощью этого метода можно изучать и сравнивать ОПАЛы отдельных сайтов. Также метод может помочь в определении эффекта мутаций у видов, для которых невозможно изучать этот вопрос экспериментально.

С помощью нового метода мы нашли для поверхностного белка вируса иммунодефицита человека 1 (ВИЧ1) аминокислоты, меняющие приспособленность между вирусными подтипами А, В и С. Результаты выложены в открытый доступ в виде веб-сервиса (http://makarich.fbb.msu.ru/galkaklink/hiv_landscape).

Таким образом, в диссертации решается важная для эволюционной биологии задача – оценка роли непостоянства однопозиционных адаптивных ландшафтов в эволюции белков.

Основные результаты и положения, выносимые на защиту

1. В митохондриальных белках видов *Opisthokonta* филогенетическое расстояние между параллельными заменами в среднем на 20% меньше, чем филогенетическое расстояние между дивергентными заменами. Это характеризует степень непостоянства однопозиционных адаптивных ландшафтов в эволюции митохондриальных белков.
2. Виды, в которых происходят замены на аминокислоты, представляющие референтные, минорные нейтральные и патогенные аллели в митохондриальных белках человека, более близко родственны ему, чем виды, в которых происходят замены на аминокислоты, не встречающиеся у человека.
3. Информация о взаимном расположении аминокислотных замен на филогенетическом дереве позволяет находить аминокислоты, меняющие приспособленность между кладами. Это может быть полезным в поиске причин различий в особенностях заражения, течения инфекции и ответа на лечение между разными штаммами одного вируса.

Объекты и методы

В работе изучены ОПАЛы кодируемых и работающих в митохондриях белков *Opisthokonta*, а также ОПАЛы поверхностных белков gp160 вируса

иммунодефицита человека (ВИЧ-1) и гемагглютини́на вируса гриппа человека А. Эти объекты выбраны, так как последовательности генов митохондриальных и вирусных белков прочитаны для тысяч видов/штаммов, что необходимо для применения наших методик изучения ОПАЛов. Кроме того, для указанных вирусных белков опубликованы экспериментально измеренные аминокислотные предпочтения в некоторых штаммах, что позволило нам сравнить наши результаты с экспериментальными.

В работе использованы как стандартные методы работы с данными секвенирования (для выравнивания последовательностей, реконструкции филогенетических деревьев и поиска положительного отбора), так и оригинальные методы анализа данных для поиска непостоянства ОПАЛов.

Апробация работы и публикации по теме диссертации

Результаты и основные положения работы доложены на международных конференциях SMBE-2014 (Пуэрто-Рико, Сан-Хуан), SMBE-2015 (Австрия, Вена), SMBE-2017 (США, Остин), SMBE-2018 (Япония, Йокогама), Всероссийской конференции ИТиС-2015 (Сочи), а также на семинаре группы эволюционной геномики австрийского Института науки и технологии (Австрия, Клостернойбург, 2018).

По теме диссертации опубликовано 2 работы в рецензируемых международных научных журналах, индексируемых в базах данных Scopus и Web of Science.

Работа по теме диссертации была поддержана грантом 18-34-00358 мол_а (руководитель Г.В.Клинк).

Структура и объём работы

Диссертация состоит из введения, 4 глав, выводов и списка литературы, состоящего из 99 источников. Содержательная часть работы изложена на 109 страницах текста, включает 25 рисунков, 2 таблицы и 14 приложений.

Содержание работы

*1) Исследование степени вариабельности однолокусных адаптивных ландшафтов митохондриальных белков *Opisthokonta**

Мы изучили роль непостоянства ОПАЛов в эволюции митохондриальных белков *Opisthokonta* с помощью оригинального филогенетического метода. Изменение частоты замен на аминокислоту в ходе эволюции может отражать изменение относительной приспособленности этой аминокислоты.

Мы сравнивали филогенетические расстояния между параллельными и дивергентными заменами, складывая информацию для всех аминокислот и всех сайтов белка. При этом, чтобы предотвратить возможные влияния объединения сайтов и аминокислот на результат (Рисунок 1), для каждой предковой аминокислоты в сайте мы выбирали максимально возможное одинаковое число пар параллельных и дивергентных замен, повторяя эту процедуру для всех предковых аминокислот во всех сайтах и получая две подвыборки параллельных и дивергентных замен одинакового размера. В обеих подвыборках для каждой пары замен мы измеряли филогенетические расстояния между этими заменами. Полученные расстояния разделяли на окна, для каждого окна считая отношение числа параллельных и дивергентных замен, попавших в это окно - Г/Д (гомоплазии/дивергентные замены). Чтобы удостовериться, что разные сайты вносят согласованный

вклад в значение статистики, мы проводили бутстрэппинг (взятие подвыборки с возвращением) сайтов в 1000 повторностях, каждый раз повторяя всю процедуру расчёта статистики. На основании полученных данных мы считали средние значения и квантили распределения нашей статистики. В качестве контроля с неизменным адаптивным ландшафтом, мы симулировали эволюцию каждого гена с помощью программы «evolver» пакета «PAML» [Yang 2007], используя модель «empirical F» с дискретным гамма-распределением скоростей эволюции между сайтами.

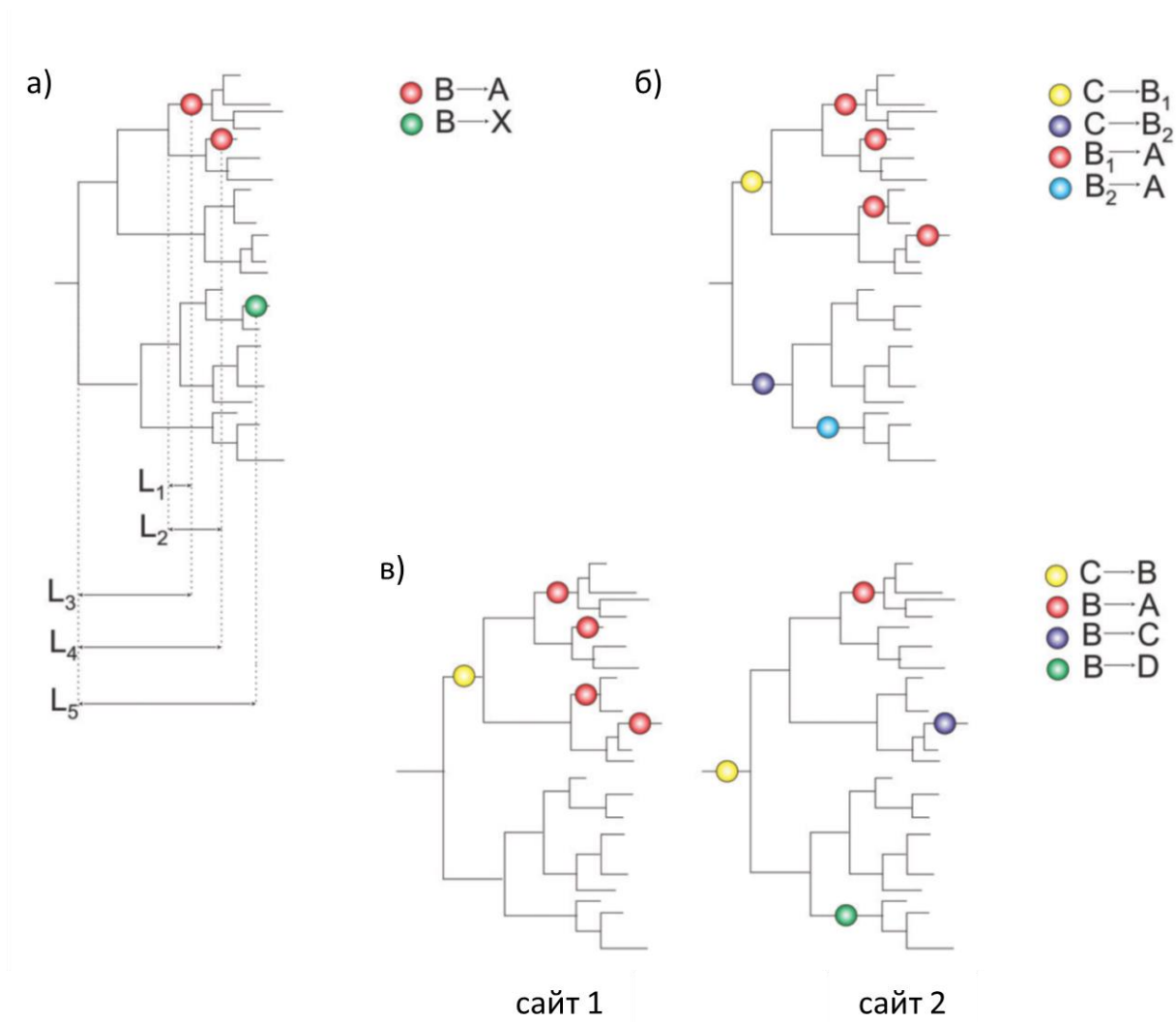


Рисунок 1. Определение филогенетических расстояний между параллельными и дивергентными заменами. Кругами отмечены аминокислотные замены, произошедшие на соответствующих ветвях филогенетического дерева.

а) – для каждой пары аминокислот (B , A) в сайте белка мы считаем филогенетические расстояния между всеми параллельными заменами $B \rightarrow A$ ($L1+L2$) и расстояния между всеми дивергентными заменами $B \rightarrow A$ и $B \rightarrow X$ ($L3+L5$, $L4+L5$), где X – любая аминокислота, кроме A и B .

б) – замена $B_1 \rightarrow A$ более частая, чем $B_2 \rightarrow A$, что при отсутствии контроля на предковую аминокислоту даст избыток гомоплазий на коротких филогенетических расстояниях.

в) – объединение информации из сайтов с разными свойствами также может приводить к избытку гомоплазий на коротких филогенетических расстояниях.

Помимо непостоянства ОПАЛов, кластеризация параллельных замен в более близких видах может наблюдаться из-за изменения мутационной матрицы, смены предпочтений кодонов, ошибок филогенетической реконструкции. С помощью дополнительных тестов мы показали, что наш результат обеспечен именно вариабельностью ОПАЛов. В частности, для проверки роли изменения мутационной матрицы мы провели симуляцию эволюции с отдельными матрицами замен для позвоночных, беспозвоночных и грибов, вычисленными на основании наших данных программой “RAxML”. Мы провели свой анализ для полученных таким образом последовательностей и получили результаты, неотличимые от результатов для симуляции, проведённой с использованием одной матрицы замен для всего дерева. Для борьбы с влиянием ошибок филогенетической реконструкции мы выполнили

отдельный анализ только для пар замен, которые определили, как «надёжные». Для выявления таких замен, для каждой ветви филогенетического дерева мы определили бутстрэп-поддержку на основании ста повторностей процедуры бутстрэппинга сайтов с помощью программы «RAxML». Пару замен считали «надёжной», если по крайней мере одна ветвь между ними имела 100% поддержку бутстрэпа, и для каждой замены из пары предковая и производная аминокислоты имели оценку максимального правдоподобия 1 по расчётам в программе «codeml». Это обеспечивало надёжность топологии и реконструкции предковых состояний для пары замен.

Мы обнаружили избыток параллельных замен на близких филогенетических расстояниях для 5 митохондриальных белков 4350 видов *Opisthokonta*. Максимальное отношение Г/Д = 2 наблюдалось между близкими видами и монотонно снижалось с увеличением филогенетического расстояния (Рисунок 2а). В симулированных данных наблюдалось незначительное снижение значения Г/Д с расстоянием, что может быть следствием несбалансированности филогении. Среднее филогенетическое расстояние между параллельными заменами в среднем оказалось на 20% меньше, чем между дивергентными заменами (Рисунок 2б).

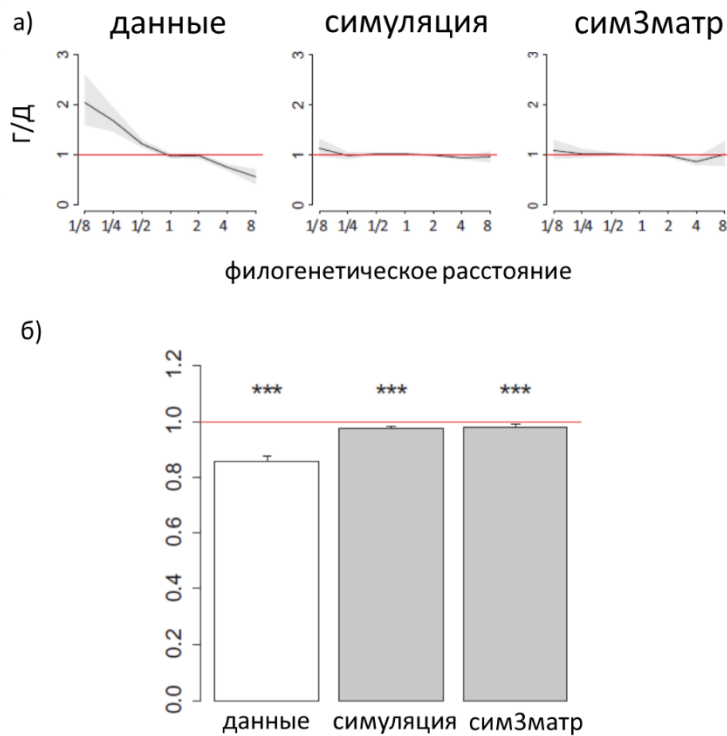


Рисунок 2. Отношение числа параллельных и дивергентных замен на разных филогенетических расстояниях (а) и отношение средних филогенетических расстояний между параллельными и дивергентными заменами (б) в филогении для 4350 видов для «надёжных» пар замен. **симЗматр** - симуляция с независимыми матрицами замен в каждой кладе.

Таким образом, непостоянство однопозиционных ландшафтов приспособленности играет заметную роль в эволюции митохондриальных белков животных. Скорость изменения ОПАЛов имеет порядок скорости аминокислотных замен.

2) *Вариабельность приспособленности аллелей в митохондриальных белках человека*

Мы проанализировали встречаемость различных аллелей митохондриальных белков человека в 4350 видах *Opisthokonta* и показали, что замены на такие аминокислоты чаще происходят в более родственных человеку линиях.

Мы использовали метод, описанный в предыдущем разделе, с той разницей, но брали замены не парами, а по одной, и сравнивали филогенетические расстояния от ветви человека до замен на аминокислоты, встречающиеся и не встречающиеся у людей как референтные, нейтральные или патогенные аллели.

Для всех трёх типов аллелей человека отношение Г/Д снижалось с филогенетическим расстоянием от человеческой ветви, и среднее филогенетическое расстояние от человека до замен на эти аминокислоты было на 10% меньше, чем расстояние до замен на аминокислоты, не встречающиеся у людей (Рисунок 3). Для патогенных аминокислот эффект наблюдался только для вариантов, встречающихся в состоянии гомоплазии (когда наблюдаемый аллель – единственный вариант в организме).

Отсутствие эффекта для аминокислот, наблюдавшихся только в гетероплазмичном (полиморфном) состоянии, может объясняться тем, что в гомоплазмичном состоянии они летальны для человека и не могут становиться хуже с филогенетическим расстоянием.

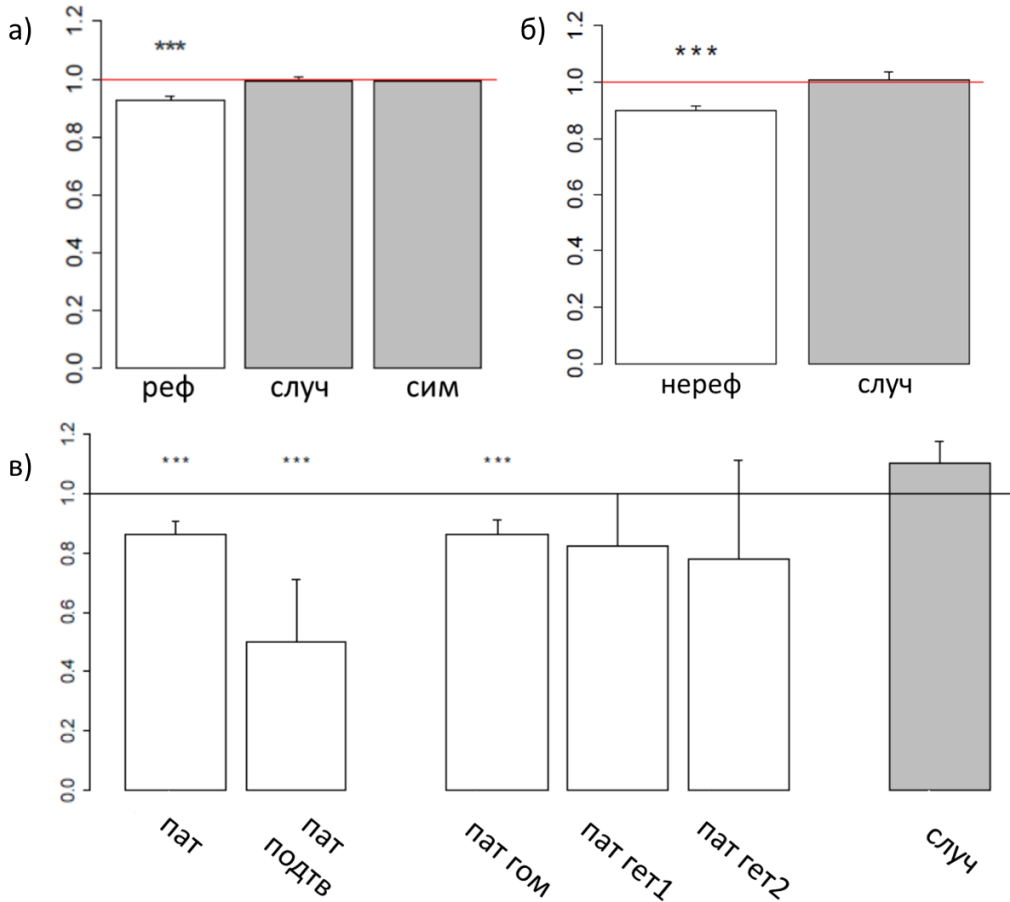


Рисунок 3. Отношение средних филогенетических расстояний между человеком и заменами на аминокислоты, референтные (а), нейтральные (б) и патогенные (в) для человека. **реф** – референтные аминокислоты, **нереф** – нейтральные нереферентные варианты, **пат** – патогенные варианты, **случ** – случайно выбранные аминокислоты, встречающиеся в тех же сайтах, **сим** – симуляция *evolver*, **пат подтв** – варианты со статусом «confirmed» в базе данных MITOMAP, **пат гом** – патогенные варианты, встречающиеся в состоянии гомоплазии, **пат гет1** – патогенные варианты, наблюдавшиеся только как гетероплазмичные, **пат гет2** – патогенные варианты, наблюдавшиеся только как гетероплазмичные, с двумя и более литературными ссылками в «MITOMAP». *** - $p < 0.001$.

Нейтральные и патогенные аллели в среднем оказались ближе по физико-химическим свойствам к референтному аллелю человека, чем аминокислоты, не встречающиеся в людях (Рисунок 4).

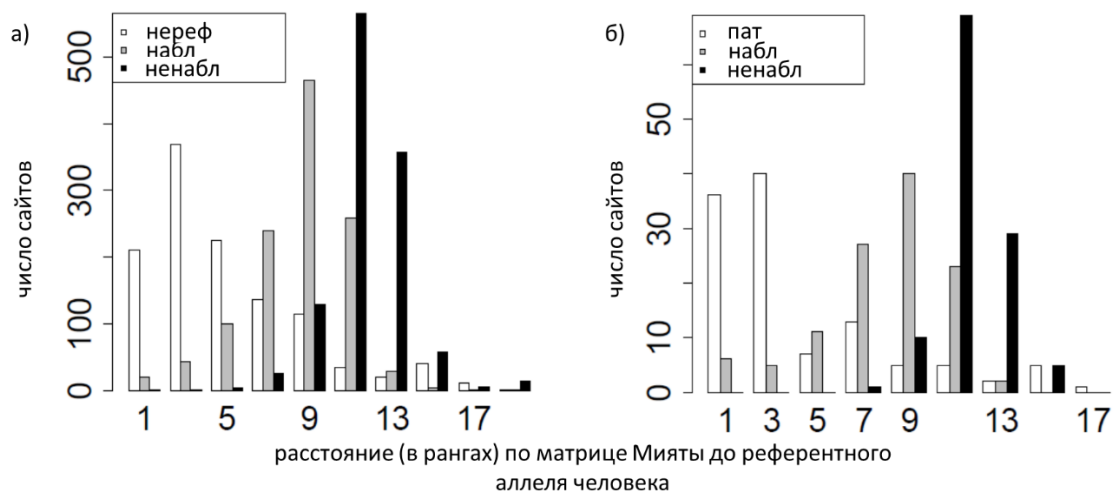


Рисунок 4. Ранг расстояния по матрице Мияты между референтным аллелем человека и его нереферентными аллелями (белый цвет), другими аминокислотными вариантами, которые наблюдались (серый) или не наблюдались (чёрный) в том же сайте:

- а) нереферентный аллель – нейтральный полиморфизм,
- б) нереферентный аллель – патогенный вариант.

Для каждого сайта с известными нейтральными или патогенными вариантами, мы проранжировали (от минимума к максимуму) все аминокислоты по расстоянию Мияты от референтного аллеля человека и определили ранг для патогенного (или нейтрального) варианта человека, средний ранг для аминокислот, встречающихся в сайте в других видах и средний ранг для остальных аминокислот.

Наши результаты подтверждают роль вариабельности адаптивных ландшафтов в эволюции митохондриальных белков и показывают, что встречаемость аминокислоты в близких к человеку видах не обязательно является свидетельством её безвредности для человека, что стоит учитывать при оценке эффекта мутаций.

3) Поиск аминокислот с вариабельной приспособленностью

Предыдущие разделы диссертации были посвящены свидетельствам вариабельности ОПАЛов, которые мы получили с применением совокупных статистик, использующих информацию из множества аминокислотных сайтов белков. Этот раздел описывает разработанный нами подход, позволяющий находить аминокислоты с разной приспособленностью в группах видов при наличии аминокислотных последовательностей и восстановленной филогении. Мы выбираем на филогенетическом дереве два фокальных вида и ищем аминокислоты, замены на которые происходят ближе к одному виду и дальше от другого, чем ожидается. Распределение ожидаемых филогенетических расстояний строится для случайных подвыборок такого же числа замен из тех же предковых аминокислот, что и замены на аминокислоту под рассмотрением. Таким образом, наш метод применим только для сайтов с числом замен, достаточным для достижения желаемого уровня значимости.

С помощью симуляций эволюции со сменой ОПАЛов между двумя кладами филогенетического дерева мы показали, что наш метод лучше всего работает для средних коэффициентов отбора и его мощность растёт с увеличением количества данных (Рисунок 5).

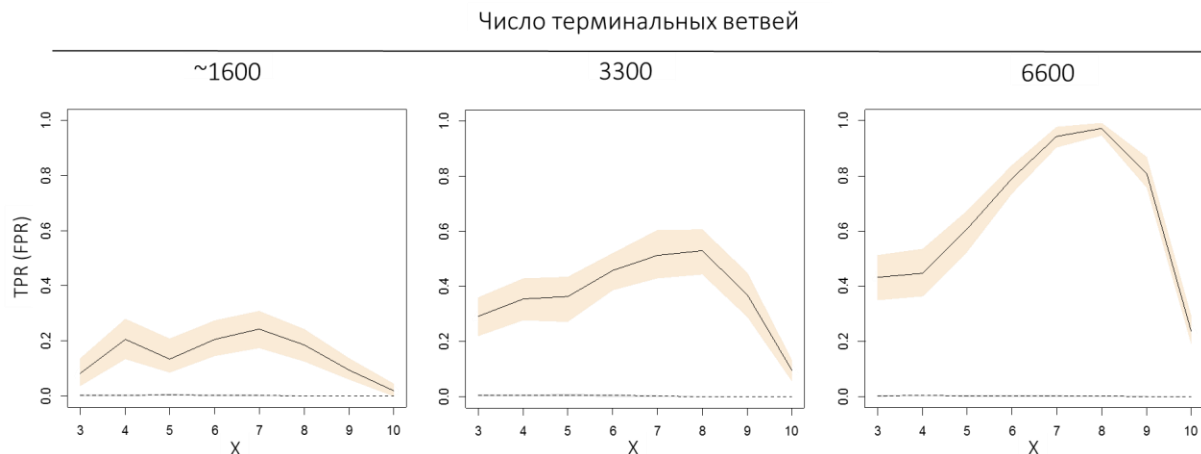


Рисунок 5. Результаты нашего подхода в симуляциях при разных значениях популяционной приспособленности ($X = Ne \times s$) предпочтительного варианта для трёх филогенетических деревьев с разным числом терминальных ветвей; сплошная линия и бежевая область – средняя доля истинно положительных классификаций (True Positive Rate, TPR) и 90% доверительный интервал для ста случайных выборок по 1000 аминокислот из результатов симуляции эволюции 500 сайтов; прерывистая линия и серая область – то же для доли ложноположительных классификаций (False Positive Rate, FPR).

Мы применили свой метод к поверхностному белку gp160 ВИЧ1 из двух вирусных подтипов А и В и к гемагглютнину вируса гриппа А из подтипов Н1 и Н3. Мы нашли 59 аминокислот с вариабельной приспособленностью (АВП) в 46 сайтах для gp160 и 9 АВП в 7 сайтах для гемагглютинаина и показали, что полученные нами результаты согласуются с результатами экспериментов по глубокому мутационному сканированию (Haddox et al. 2016, Haddox et al. 2018, Lee et al. 2018) (Рисунок 6). Для этого для каждой аминокислоты мы нашли отношение r её экспериментально измеренных приспособленностей в штамме, к которому замены на неё происходили ближе, и в штамме, к которому замены на неё происходят дальше по результатам нашего теста. Значения r для АВП значительно превышали

таковые для остальных аминокислот (односторонний тест Уилкоксона, $p = 0.005$ gp160, $p = 0.2$ для гемагглютинаина), что говорит о согласовании филогенетически и экспериментально определённых изменений аминокислотных приспособленностей.

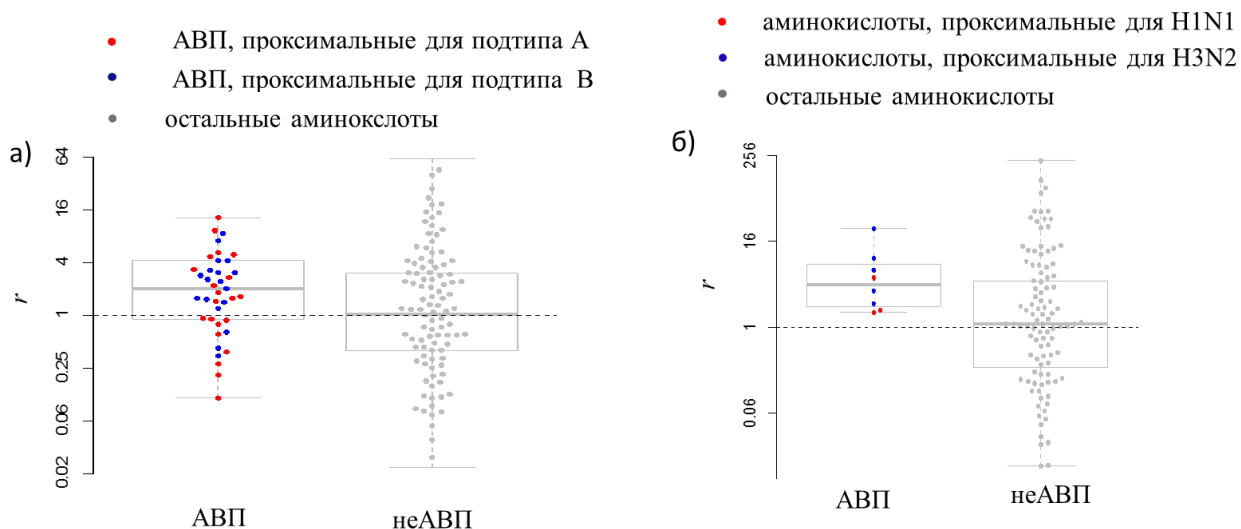


Рисунок 6. Диаграмма размаха для отношения экспериментально измеренных приспособленностей аминокислоты в штамме, к которому замены на неё происходят ближе, и в штамме, от которого замены на неё происходят дальше. Показаны аминокислоты со значимым (слева, $p < 0.01$ для обоих штаммов) или незначимым (справа, $p > 0.1$ для обоих штаммов) изменением приспособленности между штаммами; сплошная линия – медиана, границы прямоугольников обозначают межквартильный размах (IQR) между первым ($Q1$) и третьим ($Q3$) квартилями, усы обозначают интервалы $Q1 - 1.5 \times IQR$ и $Q3 + 1.5 \times IQR$, выбросы не показаны; диаграммы построены по всем данным; каждая АВП, проксимальная к подтипу А или В ВИЧ1 (а) и подтипу Н1 или Н3 (б), и подвыборка из 100 незначимых аминокислот показаны как красные, синие и серые точки соответственно. Пунктирная линия обозначает значение r при постоянной экспериментальной приспособленности ($r=1$).

Изменения приспособленности аллеля могут привести к эпизодам действия положительного отбора. Для исследования связи между этими процессами, мы нашли сайты gp160, в которых отношение скоростей несинонимичной и синонимичной эволюции (dN/dS) превышает нейтральное значение по оценке программы «codeml» ($\omega > 1$) в подтипе А, подтипе В или обоих подтипах. Среди 46 сайтов, несущих АВП, только 20 (43%) испытывали положительный отбор в подтипе А и/или В, несмотря на то, что мощность теста dN/dS и мощность нашего теста возрастают при одном и том же условии увеличения числа замен. Несмотря на избыток сайтов под положительным отбором среди сайтов с АВП (точный тест Фишера, двусторонний $P < 0.001$), отсутствие свидетельств положительного отбора более, чем в половине сайтов с АВП, говорит о том, что эти явления часто могут быть несвязанными.

Наш метод может находить АВП для нескольких фокальных видов. Мы продемонстрировали это, найдя 327 АВП, приспособленность которых меняется между подтипами ВИЧ1 А, В и С так, что в одном из подтипов она больше или меньше, чем в двух других. Списки аминокислот доступны на веб-странице http://makarich.fbb.msu.ru/galkaklink/hiv_landscape/, которая может быть полезна при изучении различий вирусных подтипов в течении заболевания и ответе на лечение.

Выводы

1. В митохондриальных белках *Opisthokonta* филогенетическое расстояние между параллельными заменами в среднем на 20% меньше расстояния между дивергентными заменами. Это является свидетельством вариабельности адаптивных ландшафтов митохондриальных белков.
2. В среднем приспособленность референтных, нейтральных и патогенных (но не летальных) аллелей митохондриальных белков человека снижается с филогенетическим расстоянием от ветви *Homo sapiens*. Таким образом, приспособленность встречающихся в популяциях человека аминокислот выше в видах, более эволюционно близких к человеку. Обнаруженный в популяциях человека патогенный вариант может требовать меньше изменений в других сайтах белка, чтобы стать разрешённым, чем вариант, не наблюдаемый в человеке.
3. Разработан и проверен на экспериментальных данных метод поиска индивидуальных аминокислот, приспособленность которых меняется между двумя участками филогенетического дерева. Найдены аминокислоты поверхностных белков ВИЧ-1 и вируса гриппа А, меняющие приспособленность между вирусными подтипами. Список таких аминокислот для поверхностного белка из подтипов А, В, С ВИЧ-1 представлен в виде веб-сервера. Эта информация может помочь в изучении отличий между вирусными подтипами, приводящих к разному течению болезни и ответу на лечение. Показано, что изменение однопозиционного адаптивного ландшафта не всегда влечёт за собой сильный положительный отбор.

Публикации в изданиях из перечня ВАК

- 1) Klink G. V., Bazykin G. A. Parallel evolution of metazoan mitochondrial proteins //Genome biology and evolution. – 2017. - P. 1341-1350
- 2) Klink G. V., Golovin A. V., Bazykin G. A. Substitutions into amino acids that are pathogenic in human mitochondrial proteins are more frequent in lineages closely related to human than in distant lineages //PeerJ. – 2017. – Vol. 5. – P. e4143.
- 3) Klink G. V., Bazykin G. A. Phylogenetic inference of changes in amino acid propensities with single-position resolution //bioRxiv. – 2019. – P. 589572.

Тезисы конференций

Klink G.V., Bazykin G.A. Inference of prevalence of epistasis from huge phylogenies// Society for Molecular Biology & Evolution (SMBE 2014), Сан-Хуан, Пуэрто-Рико, 2014

Klink G.V., Bazykin G.A. Analysis of prevalence of epistasis on the basis of huge phylogenies// Society for Molecular Biology & Evolution (SMBE 2015), Вена, Австрия, 2015

Klink G.V., Bazykin G.A. Analysis of prevalence of epistasis on the basis of huge phylogenies// Информационные технологии и системы (ИТИС 2015), Сочи, Россия, 2015.

Klink G.V., Bazykin G.A. Inference of changes of fitness landscape from sequence data with single-position resolution // Society for Molecular Biology & Evolution (SMBE 2017), Остин, США, 2017

Klink G.V., Bazykin G.A. Inference of changes of HIV-1 gp160 protein fitness landscape from sequence data with single-position resolution // Society for Molecular Biology & Evolution (SMBE 2018), Йокогама, Япония, 2018