



На правах рукописи

Леушкин Евгений Владимирович

**Анализ эволюции инсерций и делеций в последовательности ДНК,
проводимый на основе сравнения полных геномов**

03.01.09 - математическая биология, биоинформатика

Автореферат
диссертации на соискание степени
кандидата биологических наук

Москва, 2014

Работа выполнена на факультете биоинженерии и биоинформатики Московского государственного университета им. М.В. Ломоносова и в секторе № 4 молекулярной эволюции Федерального государственного бюджетного учреждения науки Институт проблем передачи информации им. А.А. Харкевича Российской академии наук.

Научный руководитель:

Базыкин Георгий Александрович
кандидат биологических наук,
заведующий сектором №4 молекулярной эволюции ИППИ РАН

Официальные оппоненты:

Лебедев Юрий Борисович
доктор биологических наук,
заведующий лабораторией сравнительной и функциональной геномики
Федерального государственного бюджетного учреждения науки Института
биорганической химии им. академиков М.М. Шемякина и Ю.А. Овчинникова
Российской академии наук

Кочетов Алексей Владимирович
доктор биологических наук,
заместитель директора по научно-организационной работе
Федерального государственного бюджетного учреждения науки Института цитологии и
генетики Сибирского отделения Российской академии наук

Ведущая организация:

Федеральное государственное бюджетное учреждение науки Институт молекулярной
биологии им. В.А. Энгельгардта Российской академии наук

Защита диссертации состоится 18 сентября 2014 года в 14:00 на заседании
диссертационного совета Д 002.077.04 при Федеральном государственном бюджетном
учреждении науки Институте проблем передачи информации им. А.А. Харкевича
Российской академии наук по адресу: 127994, г. Москва, ГСП-4, Большой Каретный
переулок, д. 19, стр. 1.

С текстом автореферата и диссертации можно ознакомиться в библиотеке Федерального
государственного бюджетного учреждения науки Института проблем передачи
информации им. А.А. Харкевича Российской академии наук, а также на сайте ИППИ РАН
по адресу <http://www.iitp.ru/ru/dissertation/1163.htm>.

Автореферат разослан "7" августа 2014 г.

Ученый секретарь диссертационного совета

доктор биологических наук, профессор



Г.И. Рожкова

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность проблемы. Методы секвенирования нового поколения кардинально сократили время и снизили стоимость секвенирования, что привело к быстрому росту числа прочитанных геномов. Происходит как секвенирование новых видов, так и пересеквенирование генотипов многих особей видов с уже секвенированными геномами с целью получения данных популяционной изменчивости. Наличие качественных полногеномных данных по полиморфизму и дивергенции сильно расширило возможности сравнительно-геномного анализа. В частности, теперь возможно с высокой точностью измерить скорости мутагенеза для мутаций даже редких типов, определить интенсивность отрицательного и положительного отбора, действующего на мутации, оценить эффект мутаций на геномное окружение на разных эволюционных временах, а также выяснить, как влияет на мутации генная конверсия.

Основным объектом изучения мы выбрали инсерции и делеции (инделы). Инделы изучены гораздо менее подробно, чем однонуклеотидные замены, однако они представляют большой интерес, так как обладают в среднем более радикальным эффектом на функцию участка генома. Основным организмом для исследования была выбрана плодовая мушка *Drosophila melanogaster*. *D. melanogaster* является хорошо изученным модельным организмом; для неё имеются высококачественные данные по секвенированию и пересеквенированию генома, геном хорошо аннотирован. В отличие от человека, *D. melanogaster* обладает высокой эффективной численностью и популяционной изменчивостью, что значительно расширяет возможности сравнительно-геномного анализа.

Цели и задачи исследования. Получить данные по полиморфизму инсерций и делеций. На основе данных по полиморфизму и дивергенции измерить скорости мутагенеза, приводящего к возникновению инделов, оценить действие отрицательного и положительного отбора. Оценить силу и продолжительность

воздействия возникшего индела на эволюцию окружающей последовательности ДНК. Определить эффект генной конверсии на эволюцию инделов.

Научная новизна и практическая значимость. Инделы изучены гораздо хуже одонуклеотидных замен из-за меньшего количества данных и большей трудоёмкости анализа. Однако увеличение объёма данных и совершенствование методик позволяет изучать эволюцию инделов более детально. В нашей работе мы впервые анализируем данные по полиморфизму и дивергенции инделов на полногеномном масштабе в модельном виде, чтобы изучить мутагенез, приводящий к возникновению инделов, и отбор, действующий на них. Использование в дополнение к этому данных по скоростям рекомбинации для разных участков генома позволяет изучать действие генной конверсии на эволюцию инделов.

Помимо изучения основных характеристик эволюции инделов, мы исследуем воздействие возникновения индела в определённой позиции кодирующей последовательности на эволюцию геномного окружения этой позиции. Благодаря большому количеству секвенированных видов возможно проследить влияние индела на разных эволюционных временах.

Публикации и апробация работы. По материалам диссертации опубликовано 3 статьи в международных рецензируемых научных журналах. Результаты работы были представлены на международных конференциях MCCMB'11, OWS'11, SMBE'12, российской конференции ИТИС'11. Апробация работы проведена 19 сентября 2013 г. на совместном семинаре лаборатории эволюционной геномики Факультета биоинженерии и биоинформатики МГУ им. М.В.Ломоносова и сектора молекулярной эволюции Института проблем передачи информации РАН им. А.А. Харкевича.

Структура и объем работы. Диссертация изложена на 88 страницах машинописного текста и содержит следующие разделы: введение, материалы и методы, результаты в трёх главах и выводы. Диссертация включает 28 рисунков, 1 таблицу и список литературы, содержащий 90 ссылок.

СОДЕРЖАНИЕ РАБОТЫ

1) Оценка относительных скоростей мутагенеза инделов.

Для оценки скоростей мутагенеза мы использовали полиморфные мутации, которые встречаются в 1–4 генотипах в популяции, считая, что отбор не влияет на встречаемость низкочастотных мутаций в популяции (Рисунок 1). Сравнение встречаемости низкочастотных мутаций в межгенных интервалах показало, что частоты инсерций и делеций составляют соответственно $\sim 0,037$ и $\sim 0,095$ от частоты нейтральных однонуклеотидных замен, а их средние длины – 3,46 нт для инсерций и 5,08 нт для делеций. Таким образом, данные по межгенным интервалам показывают, что в отсутствие отбора инсерции с длинами 1–60 нт давали бы прибавку в 0,13 нт в расчёте на одну нейтральную замену; в то же время делеции сокращали бы геном на 0,48 нт, что в сумме при отсутствии отбора приводило бы к потере 0,35 нт на каждую нейтральную замену.

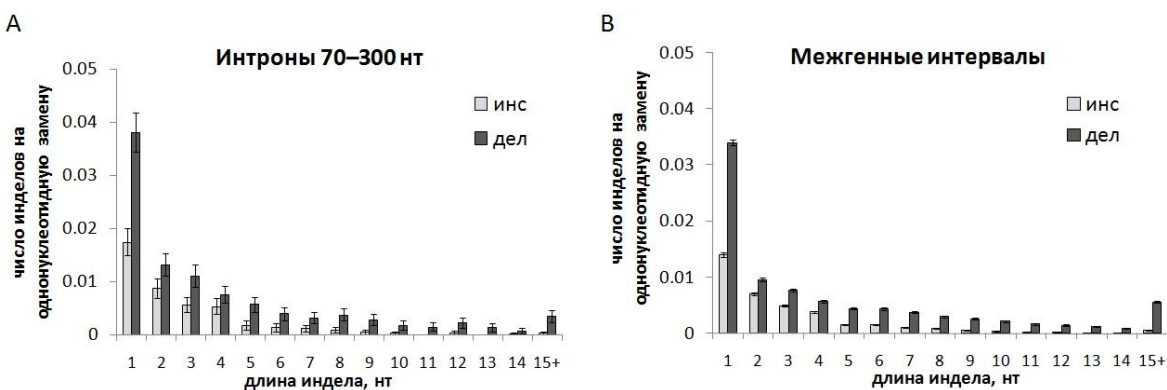


Рисунок 1. Количество инсерций (инс) и делеций (дел) разных длин в расчёте на однонуклеотидную замену для мутаций с очень низкими частотами (1–4 генотипа из 162) в коротких интронах (А) и в межгенных интервалах (В).

2) Отрицательный и положительный отбор на инделы в различных участках генома.

Сильный отрицательный отбор приводит к недопредставленности даже редких мутаций. Поэтому мы можем оценить сильный отрицательный отбор, сравнивая количества низкочастотных полиморфизмов в различных участках генома (Рисунок 2). Низкочастотные инделы обладают практически одинаковой встречаемостью в интронах различных длин и в межгенных интервалах, что свидетельствует об отсутствии сильного отбора на *de novo* мутации (Рисунок 2B-D, светло- и тёмно-серые столбики). Напротив, в экзонах даже низкочастотные инделы встречаются значительно реже (Рисунок 2A). Количество инделов снижено в экзонах более чем в 2 раза по сравнению с остальными участками генома. Встречаемость несинонимичных замен также снижена и составляет ~ 0.35 от числа однонуклеотидных замен в данном компартменте.

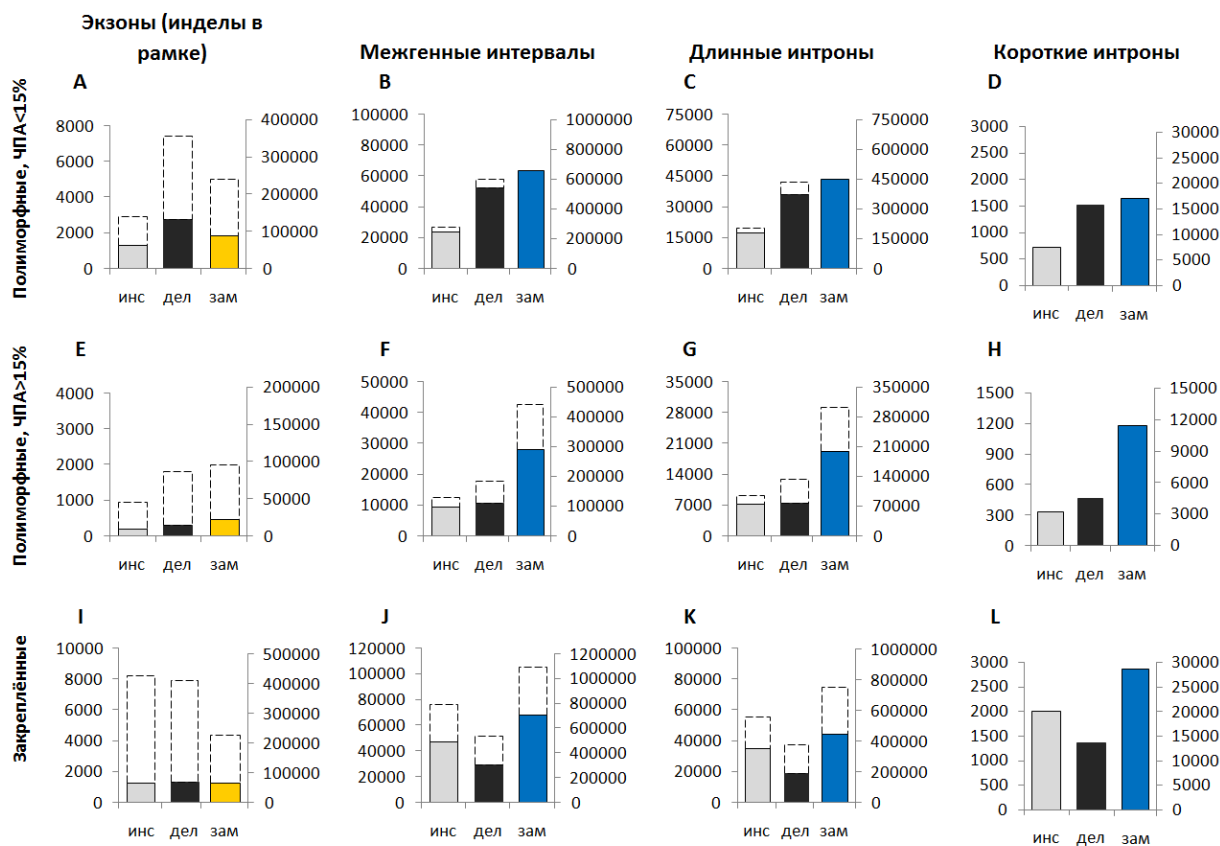


Рисунок 2. Число полиморфных и закреплённых инделов и однонуклеотидных замен в различных участках генома. Верхний ряд (А–D) – полиморфизмы с частотой производного аллеля (ЧПА) <15%; средний ряд (Е–Н), полиморфизмы с ЧПА >15%; нижний ряд (I–L), фиксированные мутации. А, Е, I: кратные трём инделы и миссенс-замены в экзонах; В, F, J: мутации в межгенных интервал; С, G, K: мутации в длинных (>300 нт) интронах; D, H, L: мутации в коротких (70–300 нт) интронах. Светло-серый соответствует инсерциям; тёмно-серый соответствует делециям; жёлтый соответствует миссенс-заменам; синий, однонуклеотидным заменам в некодирующих участках. На каждой панели левая вертикальная ось показывает количество инделов, правая вертикальная ось показывает количество однонуклеотидных замен.

Действие слабого отрицательного отбора ($-1 > N_e s > -5$) можно оценить из анализа спектра аллельных частот. Как оценку доли нейтральных мутаций мы используем ξ , отношение высокочастотных полиморфизмов к низкочастотным для потенциально отбираемых мутаций, нормированное на то же отношение, но для нейтрального класса мутаций.

Сравнение ξ для разных типов мутаций показывает, что в каждом участке генома инделов вреднее однонуклеотидных замен; при этом делеции вреднее инсерций (Рисунок 3А). Доля вредных мутаций среди полиморфизмов является наибольшей в экзонах: по крайней мере ~79% инсерций и ~88% делеций являются вредными; для сравнения, доля вредных мутаций среди однонуклеотидных миссенс-замен составляет ~72%. В межгенных интервалах и в длинных интронах доля вредных мутаций среди инсерций равна ~71%, что меньше, чем для делеций (~82%), но много больше, чем для однонуклеотидных замен (~49%) (Рисунок 3А). Отбор оказался наиболее слабым для всех типов мутаций в коротких интронах (Рисунок 3А); однако даже в этом участке генома инделов гораздо чаще являются

вредными (в ~68% случаев для инсерций и ~73% случаев для делеций), чем однонуклеотидные замены (в ~23% случаев).

Очень короткие интроны (с длинами <60 нуклеотидов) заслуживают отдельного рассмотрения: ограничение на минимальную длину интрона ведёт здесь к дополнительному отбору на делеции. В самом деле, зависимость отбора против делеций от длины интрона не является монотонной: доля вредных делеций выше в длинных и в очень коротких интронах по сравнению с интронами средних длин. Она является наибольшей (~81%) для делеций, которые делают интрон короче 50 нуклеотидов. Напротив, доля вредных инсерций является наименьшей в этом классе интронов (Рисунок 3В).

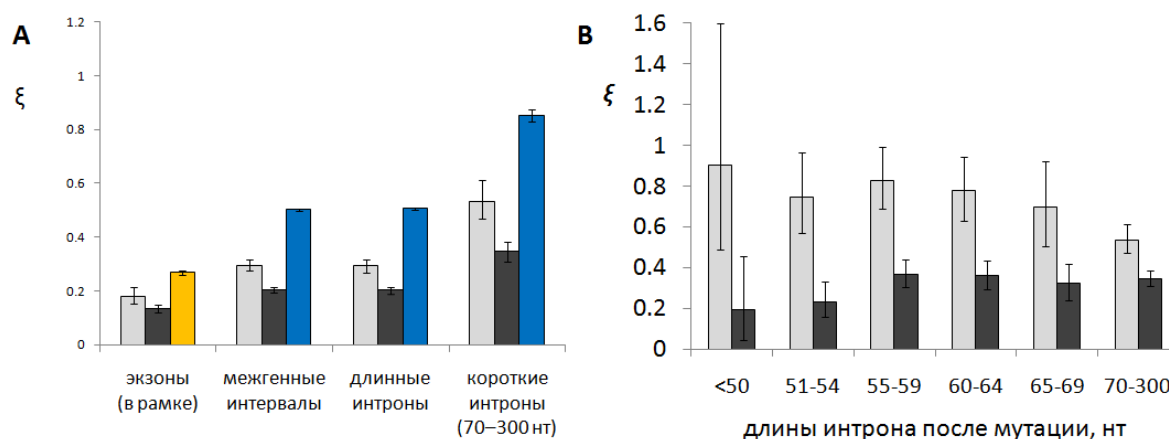


Рисунок 3. ξ для инделов и однонуклеотидных замен. Низкие значения ξ соответствуют высокой доле вредных мутаций и наоборот. (А), средние значения ξ в различных участках генома: экзонах (кратные трём инделы и миссенс-замены), межгенные интервалы, длинные (>300 нт) и короткие (70–300 нт) интроны. (В), ξ для инделов в коротких и очень коротких интронах. Светло-серым показаны инсерции; темно-серым – делеции; жёлтым, миссенс-замены; синим, однонуклеотидные замены в некодирующих участках генома.

Доверительные интервалы построены по 1000 бутстрепам.

Рисунок 4 показывает, что положительный отбор на инделы распространён повсеместно по геному *D. melanogaster*. Доля α инсерций, фиксированных под действием положительного отбора, очень велика во всех участках генома: она достигает ~61% в экзонах и ~67% в межгенных интервалах. Для делеций α ~48% в экзонах, но ниже в других участках генома: ~32–36%. Оценки α для однонуклеотидных замен значительно ниже: ~15% миссенс-замен, и ~0% замен для однонуклеотидных замен в некодирующих участках.

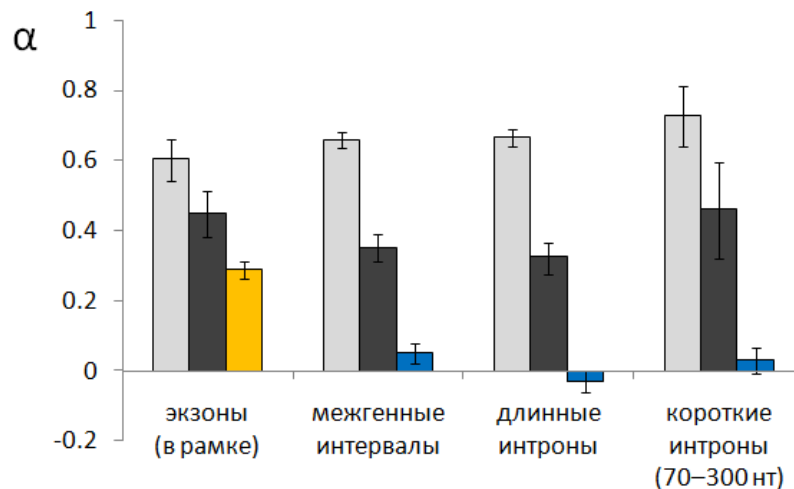


Рисунок 4. α для инделов и однонуклеотидных замен в различных участках генома: экзоны (кратные трём инделы и миссенс-замены), межгенные интервалы, длинные (>300 нт) и короткие (70–300 нт) интроны. (В), ξ для инделов в коротких и очень коротких интронах. Светло-серым показаны инсерции; темно-серым – делеции; жёлтым – миссенс-замены; синим – однонуклеотидные замены в некодирующих участках генома.

Доверительные интервалы построены по 1000 бутстрепам.

3) Влияние генной конверсии на закрепление мутаций

Чтобы изучить эффект рекомбинации на мутации различных типов, мы проанализировали эволюцию в линии *D. melanogaster*, сравнивая участки с разными скоростями рекомбинации. Сперва мы рассмотрели однонуклеотидные

замены. Скорость их фиксации – самая высокая в участках с самой низкой рекомбинацией, и в ~ 2 раза ниже в участках с высокой рекомбинацией (Рисунок 5А, левая часть). Эта зависимость, по всей видимости, возникает из-за фиксации слабовредных мутаций в участках с низкой скоростью рекомбинации. Для участков с низким отбором, а также для позиций 8–30 в интронах с длинами < 65 нт (считаются участками с самым отбором в геноме), данного эффекта не наблюдается.

Сходная картина наблюдается для инделов длин 5–10 нт, закрепившихся в популяции: как инсерции, так и делеции легче закрепляются в участках с более низкой рекомбинацией. Зависимость от рекомбинации одинакова для инсерций и делеций, в результате чего отношение инсерций к делециям от скорости рекомбинации не зависит. Совершенно по-другому зависит от рекомбинации отношение инсерций к делециям для коротких (1–4 нт) инделов. Для делеций длины 1–4 нт вероятность закрепления в популяции по-прежнему отрицательно скоррелирована со скоростью рекомбинации, причём эта корреляция сильнее, чем для более длинных делеций. В то же время для инсерций длин 1–4 нт отрицательная корреляция ослабевает с уменьшением длины инсерции и становится положительной для инсерций 1–3 нт. В результате мы наблюдаем, что отношение инсерций к делециям значительно возрастает с ростом рекомбинации (Рисунок 5В, левая часть). Самый большой контраст (в 5 раз) между участками высокой и низкой рекомбинации наблюдается для однонуклеотидных инделов; он снижается с увеличением длины, и практически отсутствует для инделов длиннее 5 нт (Рисунок 5В, левая часть).

В отличие от закрепившихся инделов, частота полиморфных инделов имеет приблизительно одинаковую зависимость от скорости рекомбинации для инсерций и делеций, так что отношение инсерций к делециям практически не зависит от скорости рекомбинации (Рисунок 5В, средняя часть). Таким образом, различие в скоростях мутагенеза не может объяснить зависимость отношения частот инсерций и делеций от рекомбинации, наблюдаемую на данных по дивергенции.

Следовательно, эта зависимость возникает в процессе закрепления мутаций. На рисунке 5В, правой части для мутаций, расщепляющихся на более высоких частотах ($>15\%$), мы видим паттерн, промежуточный между тем, что наблюдали для *de novo* мутаций и для закрепившихся мутаций.

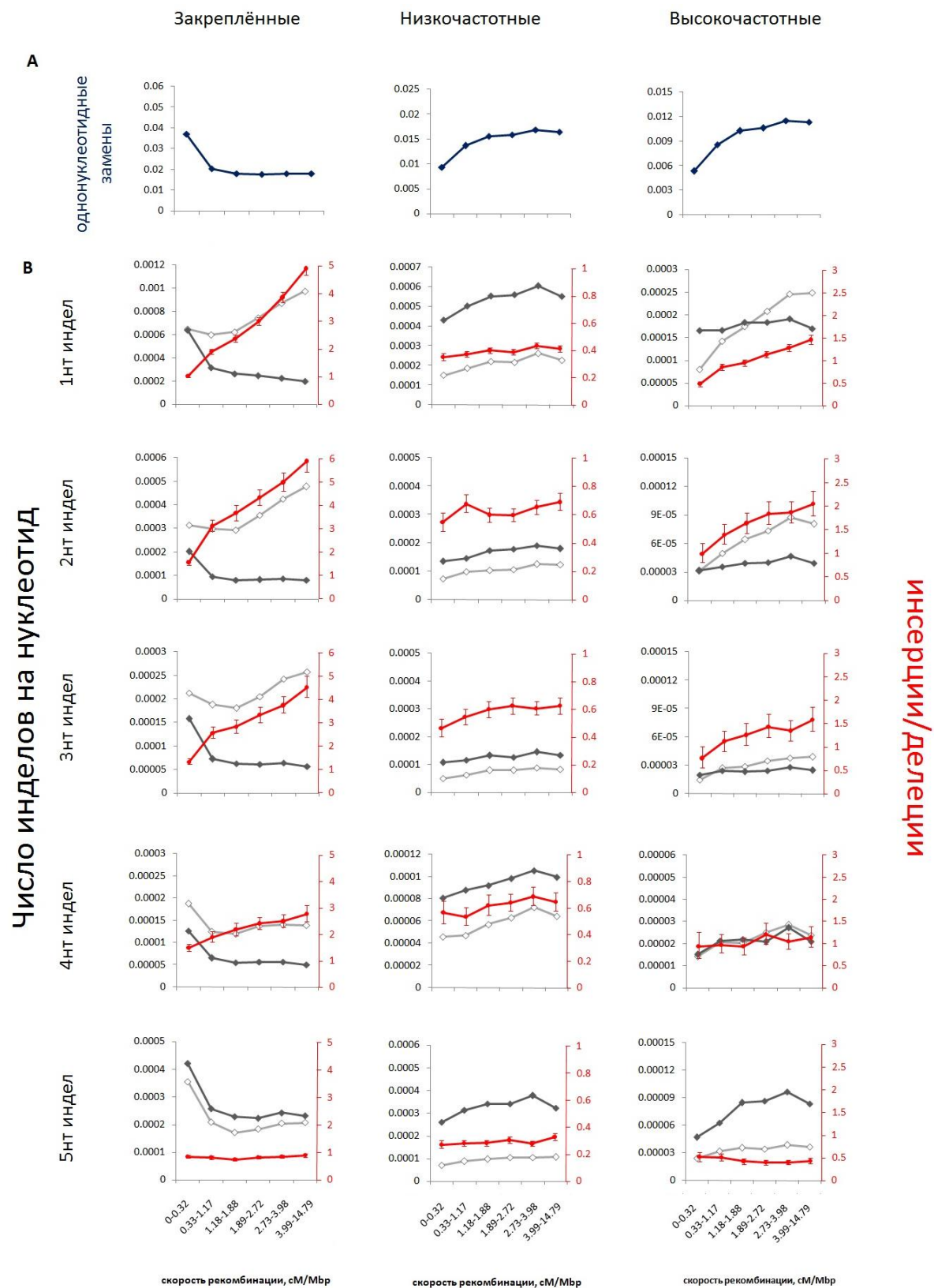


Рисунок 5. Закрепившиеся и полиморфные мутации в некодирующих участках *D. melanogaster* в различными скоростями рекомбинации. (А) Однонуклеотидные мутации. (В) Инсерции и делеции. Серым показаны инсерции, чёрным – делеции, красным – отношение инсерций к делециям, синим – однонуклеотидные замены. Левая часть – закрепившиеся мутации, средняя часть – низкочастотные мутации (<3%), правая часть – мутации с частотой выше 15%.

В процессе закрепления две направленные силы могут влиять на частоты аллеля: естественный отбор и смещённая генная конверсия. Отбор действует на аллели, различающиеся по приспособленности; положительный отбор увеличивает, а отрицательный отбор уменьшает вероятность закрепления нового аллеля. Для всех типов мутаций отрицательный отбор гораздо распространённее положительного. В отличие от отбора, смещённая генная конверсия действует невзирая на приспособленности аллелей. Действуя на пару антагонистических мутаций, таких как $A \rightarrow G$ против $G \rightarrow A$, или, в нашем случае, инсерции и делеции, она пропорционально увеличивает частоту одних и уменьшает частоту других.

Ряд наблюдений доказывает, что естественный отбор не может объяснить наблюдаемые зависимости отношения инсерций к делециям от рекомбинации. Во-первых, отношение инсерций к делециям не зависит от длины индела в участках с самой низкой рекомбинацией ($\rho < 0,32$ сМ/Мbp), где действие конверсии предполагается минимальным, в то время как вероятности закрепления как инсерций, так и делеций значительно ниже, чем для нейтральных мутаций, что говорит о том, что отбор здесь сохраняется (Рисунок 6А). Другими словами, в отсутствие рекомбинации отбор не меняет распределение длин инделов. В участках же высокой рекомбинации (Рисунок 6В-Е) наблюдается корреляция вероятности закрепления с длиной индела, отрицательная для инсерций и положительная для делеций. В этих участках длинные делеции имеют более высокую вероятность закрепления в сравнении с короткими делециями, что несовместимо с гипотезой об

отборе. Во-вторых, отбор очень слаб в коротких интронах; тем не менее, если рассматривать отдельно интроны с длиной <70 нт, наблюдаются те же самые значения положительной корреляции, что и для межгенных интервалов (Рисунок 7). В-третьих, отбор должен быть сильнее в более консервативных участках. Однако отношение числа инсерций к числу делеций зависело от консервативности участка (по значению phastCons) лишь незначительно (Рисунок 8). В то же время, для каждого диапазона консервативности наблюдалось сильное различие между участками с низкой и высокой скоростью рекомбинации: в 5 раз для инделов длины 1 нт, в 2–4 раза для инделов длины 2–4 нт (Рисунок 8). Таким образом, рекомбинация, но не степень консервативности, является главным фактором, влияющим на отношение числа инсерций к числу делеций.

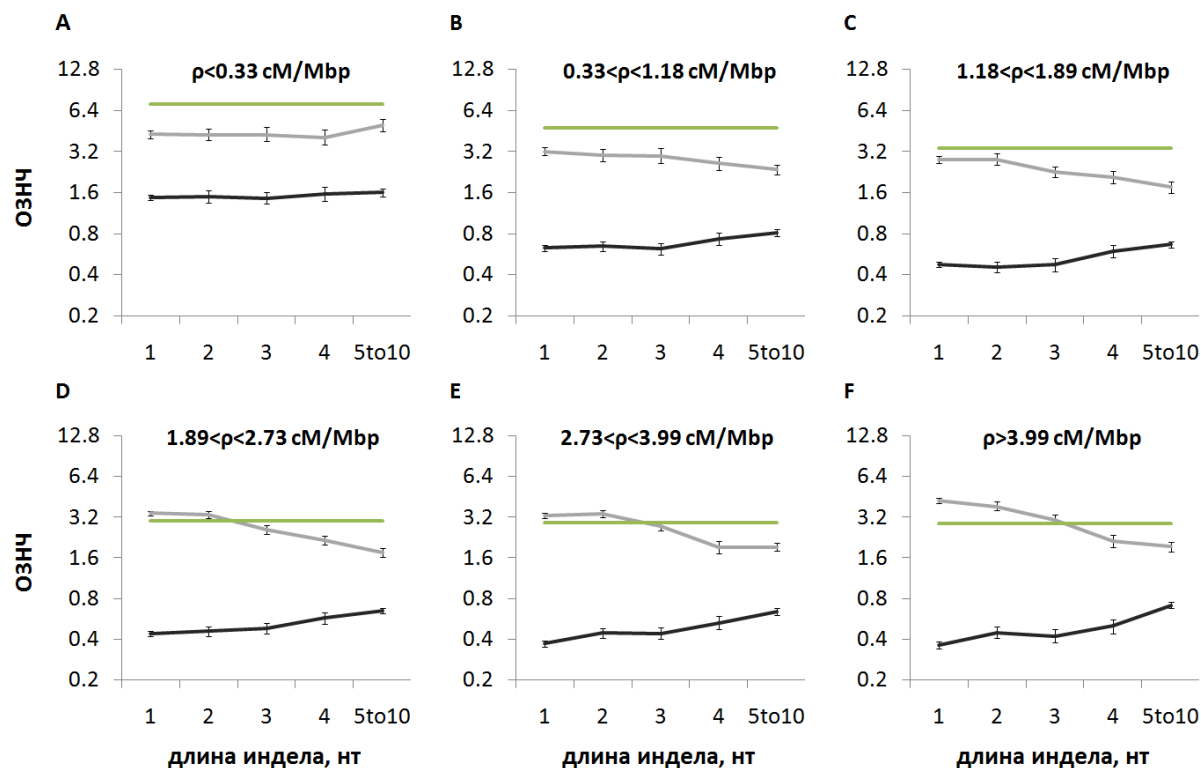


Рисунок 6. Отношение закреплённых мутаций к низкочастотным (ОЗНЧ) для участков с разными скоростями рекомбинации. Зелёным показан нейтральный стандарт, серым инсерции, чёрным делеции.

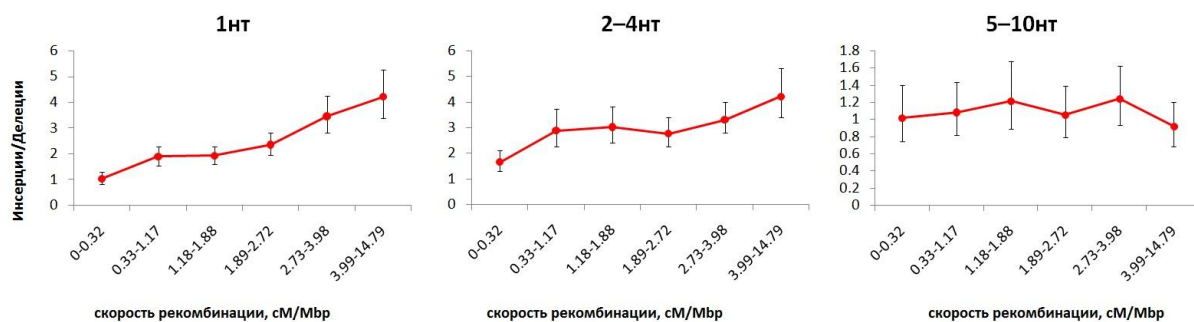


Рисунок 7. Отношение инсерции/делеции для закреплённых инделов в коротких (<75 нт) интронах *D. melanogaster*. 95% доверительные интервалы построены по 1000 бутстреп-репликам.

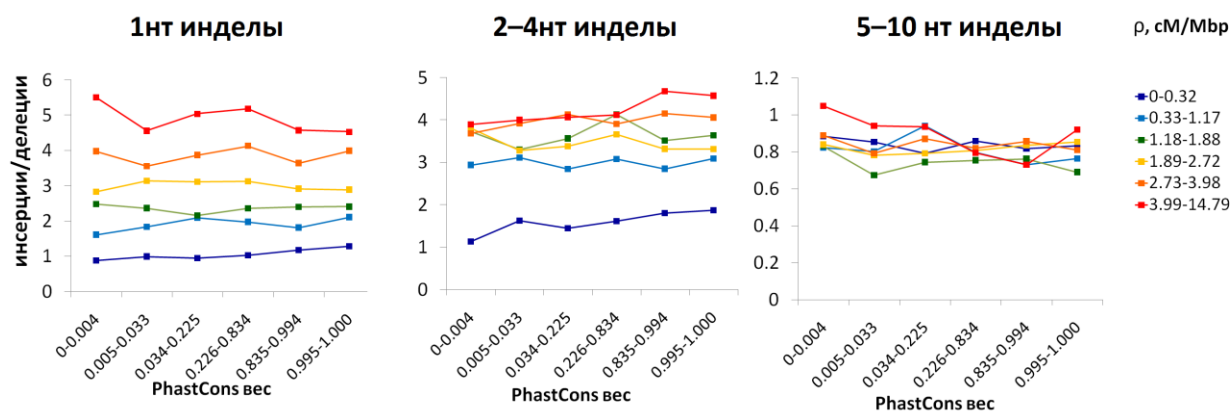


Рисунок 8. Отношение числа закреплённых инсерций к числу закреплённых делеций для некодирующих последовательностей *D. melanogaster* с различной степенью консервативности (phastCons вес) и разными скоростями рекомбинации (ρ).

Единственным правдоподобным объяснением наблюдаемого паттерна является геновая конверсия, смещённая в сторону инсерций. Как увеличение

вероятности закрепления инсерций длины 1–4 нт, так и снижение вероятности закрепления делеций длины 1–4 нт, объясняются конверсией, благоприятствующей закреплению инсерций и препятствующей закреплению делеций.

Чтобы оценить силу смещённой в сторону инсерций генной конверсии, мы сравнили частоты закрепления коротких инделов (1–4 нт) и инделов с длинами >4 нт, подразумевая, что конверсионное смещение незначимо для последних. Этот анализ показывает, что в среднем по геному генная конверсия увеличивает вероятность закрепления инсерции в 1,60, 1,55, 1,31 и 1,05 раз, и уменьшает вероятность закрепления делеции в 1,43, 1,34, 1,35 и 1,16 раз для инделов с длинами 1, 2, 3 и 4 нт соответственно. В участках с наибольшей скоростью рекомбинации эффект смещённой в сторону инсерций генной конверсии был наибольшим (Рисунок 6F), достигая увеличения в ~2 раза для однонуклеотидных инсерций и уменьшения в ~2 раза для однонуклеотидных делеций. Смещение вероятности закрепления соответствует усреднённому по геному конверсионному преимуществу для инсерций $N_e\omega = \sim 0,3-0,4$; здесь N_e – эффективный размер популяции, а $\omega = 2\kappa(\beta-0,5)$, где $(\beta-0,5)$ величина смещения конверсии, и κ – вероятность, что нуклеотидный сайт подвергнется рекомбинации. Отсутствие признаков смещённой в сторону инсерций генной конверсии в участках с низкой скоростью рекомбинации (Рисунок 6A) говорит о том, что рекомбинация является основным источником конверсии. Таким образом, κ можно рассчитать как $\kappa = \rho l x$, где ρ – скорость рекомбинации, l – длина конверсионного тракта, x – частота конверсии относительно частоты рекомбинации. В *D. melanogaster* ρ находится в пределах от 10^{-10} (низкая скорость рекомбинации) до 10^{-7} (высокая скорость рекомбинации) на нуклеотид на поколение, а $l \approx 350$ нт. Предполагая, что конверсия происходит с той же частотой, что и рекомбинация ($x = 1$), в участках с высокой скоростью рекомбинации $\kappa \approx 10^{-5}$. Наконец, $N_e = 10^5$ для *D. melanogaster*, что означает, что смещение генной конверсии в сторону инсерций $\beta = 0,7$.

4). Изменение адаптивного ландшафта при возникновении инсерций и делеций.

Логично ожидать, что такое радикальное событие в последовательности ДНК, как инсерция или делеция, может существенно повлиять на эволюцию окружающей последовательности – например, привести к изменению скорости мутагенеза или/и интенсивности отбора в данном участке. Так, можно ожидать ослабления отрицательного отбора вследствие нарушения структуры и функции участка, содержащего индел, и, как результат, увеличения числа нейтральных замен. Или же, напротив, после индела могут происходить замены, интегрирующие его в структуру белка; в таком случае мы ожидаем увидеть увеличение числа адаптивных замен.

Чтобы исследовать этот вопрос, мы использовали сравнительный анализ ортологичных белков нескольких видов *Drosophila* в целях изучения возможных эпизодов положительного отбора – адаптивных прогулок, инициируемых инделами (Рисунок 9). Мы применили тест МакДональда-Крейтмана (МК-тест) для аминокислотных замен, которые случилось на расстоянии до 100 аминокислот от сайта индела в линии *D. melanogaster* после ответвления от линии *D. sechellia*, соотнеся данные по дивергенции между *D. melanogaster* и *D. sechellia* с данными по полиморфизму для *D. melanogaster* в синонимичных и несинонимичных сайтах (Рисунок 10). Мы сравнили результаты теста для случаев, когда индел произошёл в ветке *D. melanogaster* (исследуемый случай), с соответствующими им случаями, когда индел произошёл в сестринской ветке (контроль; Рисунок 9 a vs a', b vs. b' и т.д.). Очевидно, что только в исследуемом случае, а не в контроле, аминокислотные замены могли быть вызваны инделом.

Рисунок 11 представляет данные по эволюции и положительному отбору в терминальном (т.е. после отщепления от линии *D. sechellia*) сегменте линии *D. melanogaster* для сайтов, соседствующих с инделом. Мы видим, что инделы чаще происходят в быстроэволюционирующих участках белков, как видно из того, что в аминокислотных сайтах, наиболее близких к сайту индела, число замен выше, по сравнению с более удалёнными сайтами. Это верно как для исследуемого случая, так и для контроля (Рисунок 11, верхний ряд). Однако анализ нуклеотидной

последовательности показывает, что такое ускорение не связано с повышенным мутагенезом вблизи сайта индела. На это указывает тот факт, что число синонимичных замен остаётся постоянным, при том что число несинонимичных замен увеличивается (Рисунок 11, средний и нижний ряды).

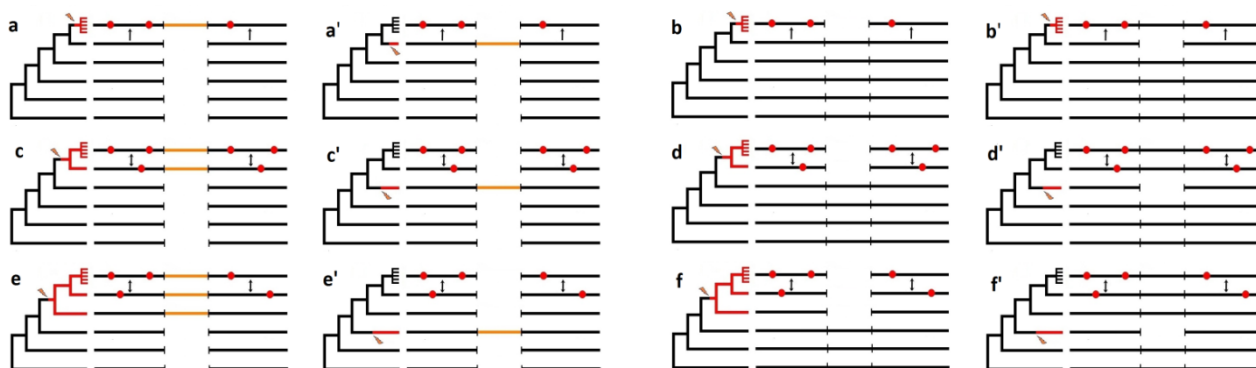


Рисунок 9. Взаимосвязь между аминокислотными заменами и инделами в белках дрозофил. На каждой части рисунка слева изображена филогения (((((D. melanogaster, D. sechellia), D. erecta), D. ananassae), D.pseudoobscura), D.virilis); время индела отмечено молнией, а сегмент эволюционного дерева, содержащий индел, выделен красным. Для D. melanogaster имеются данные по полиморфизму (обозначено гребёнкой). На рисунках а, а', с, с', е, е' изображены инсерции, а на рисунках b, b', d, d', f, f' – делеции. На рисунках а – b' (нижний ряд) представлены инделы, которые произошли до ответвления D. melanogaster от D. sechellia, а на рисунках с – d' – инделы, которые произошли на участке дерева между ответвлениями D. erecta – (D. melanogaster – D. sechellia) и D. melanogaster – D. sechellia; наконец, рисунки е – f' представляют инделы, которые произошли на участке дерева между D. ananassae – ((D. melanogaster, D. sechellia), D. erecta) и D. erecta – (D. melanogaster – D. sechellia). На рисунках а, b, c, d, e, f представлены инделы, ведущие к линии D. melanogaster, а на рисунках а', b', c', d', e', f' – инделы, произошедшие в боковых ветвях. Стрелка с двумя наконечниками обозначает 2 сравниваемых участка,

использованных для подсчёта числа замен; стрелка с одним наконечником обозначает участок, на котором замены считались с помощью поляризации.

Красными точками обозначены нуклеотидные замены.

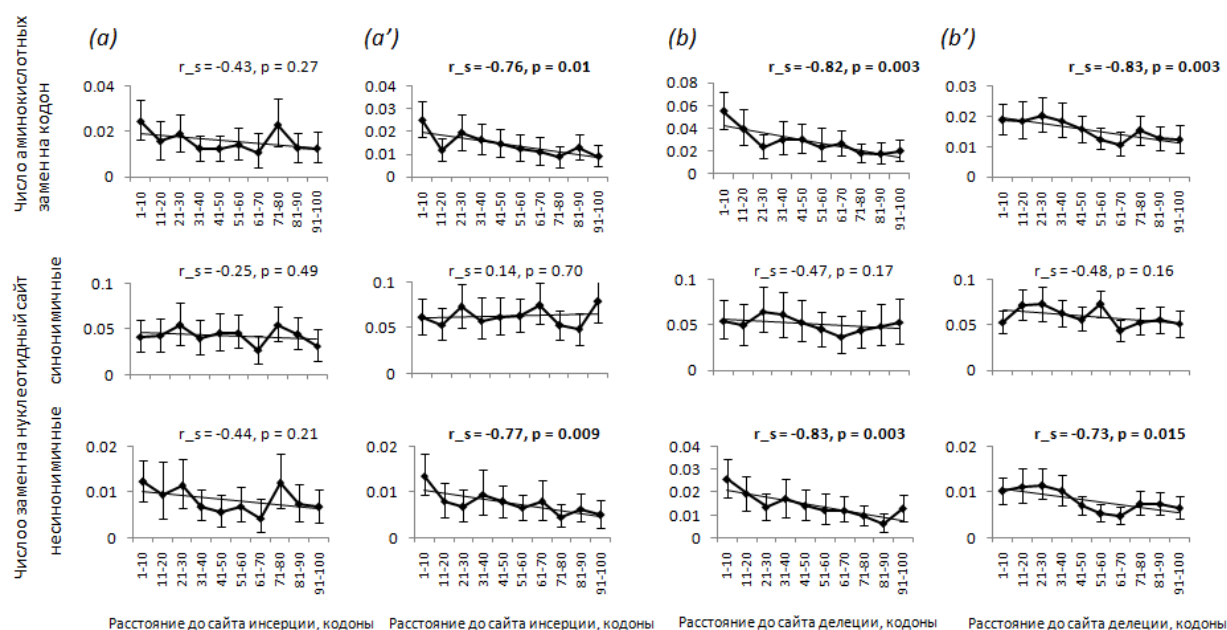


Рисунок 10. Ускоренная эволюция в несинонимичных, но не в синонимичных сайтах вблизи индела. Рисунки а–b' на данном рисунке соответствуют рисункам а–b' на рисунке 2. Верхний ряд показывает число аминокислотный замещений на аминокислотный сайт, средний ряд показывает число синонимичных замен на синонимичный сайт в ДНК, нижний ряд показывает число несинонимичных замен на несинонимичный сайт в ДНК. 95% доверительные интервалы посчитаны бутстреп-анализом с 1000 испытаний. Корреляция числа замен с расстоянием от индела проанализирована тестом Спирмана; жирным выделены случаи, когда корреляция была значимой ($p < 0,05$).

Ускоренная эволюция, связанная с инделом, также не является следствием повышения мутагенеза. На это указывает отсутствие различий в дивергенции и полиморфизме синонимичных сайтов между исследуемым случаем и контролем (Рисунок 10). Также ускорение эволюции не есть следствие ослабления отрицательного отбора, так как частоты несинонимичного полиморфизма не повышены в исследуемом случае. Лишь число несинонимичных замен было существенно повышено в соседствующих с инделом сайтах для исследуемых случаев при сравнении с контролем (Рисунок 10). Увеличение скорости эволюции было значимым для филогенетических конфигураций, представленных на рисунке 9с (инсерции) и рисунке 9b,d (делеции). В целом, после события инсерции в среднем происходит $1,03 \pm 0,75$ дополнительных аминокислотных замен в 100 аминокислотных сайтах слева и справа от позиции индела, а после события делеции – $4,77 \pm 1,03$ дополнительных замен.

Тот факт, что различие в скоростях аминокислотной эволюции обусловлено только несинонимичными заменами, означает, что все дополнительные аминокислотные замены (Рисунок 10) произошли благодаря положительному отбору.

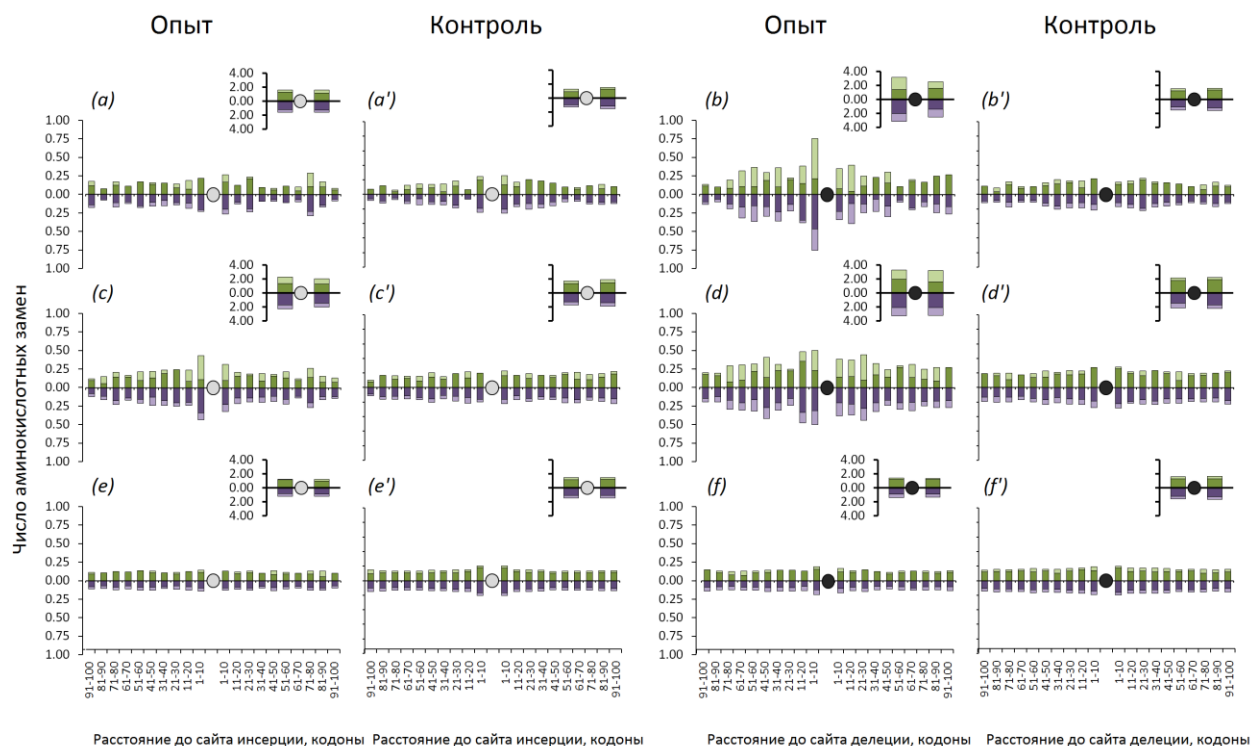


Рисунок 11. Ускорение адаптивной эволюции в аминокислотных сайтах под действием выпадений и вставок. Части рисунка соответствуют частям рисунку 9. Серым кружком обозначен сайт инсерции, чёрным кружком – сайт делеции. Слева от сайта индела – N-конец белка, справа – С-конец. Высота столбца показывает общее количество аминокислотных замен на данном участке, произошедших на терминальном сегменте линии *D. melanogaster*. Над горизонтальной осью светло-зелёным обозначена доля замен, закреплённых под действием положительного отбора, тёмно-зелёным – доля остальных замен. Под осью светло-лиловым обозначена доля замен, которые произошли в абсолютно консервативных сайтах. На каждой панели в правом верхнем углу показана суммарная информация по 100 кодоном слева и справа от сайта индела.

Наблюдаемое ускорение адаптивной эволюции зависит от времени, прошедшего с момента возникновения мутации. Наш анализ позволяет сравнивать эволюцию для инделов, произошедших сравнительно недавно, в терминальном

сегменте линии *D. melanogaster* или до ответвления от *D. erecta*, с эволюцией инделов, которые произошли до ответвления линии *D. melanogaster* от линий *D. ananassae* и *D. erecta* (Рисунок 9e–f'). Такие древние инделы уже не вызывают ускорения эволюции в терминальном сегменте (Рисунок 11 e–f'), что означает, что адаптивная прогулка продолжается в течение не очень длительного времени.

Свойства адаптивных прогулок, вызванных инсерциями и делециями, существенно различаются. По сравнению с инсерцией, делеция вызывает замены в более широком участке белка (до 100 аминокислотных остатков, по сравнению с ~40 аминокислотными остатками для инсерций); большая часть замен на всём участке закреплена под действием положительного отбора (Рисунок 11 a против b, c против d). Для делеций (Рисунок 11b-b'), но не для инсерций (Рисунок 11 a-a'), значимое увеличение числа аминокислотных замен наблюдалось даже когда индел случился в терминальном сегменте линии *D. melanogaster*, где более слабый эффект ожидался по причине того, что в такой конфигурации невозможно различить замены, которые произошли после индела, от тех, что произошли до индела.

Выводы

- 1) На 1 нейтральную однонуклеотидную замену у *D. melanogaster* в среднем происходит 0,036–0,039 инсерций (средняя длина 3,23 нт) и 0,085–0,092 делеций (средняя длина 4,78 нт). Таким образом, на уровне мутагенеза наблюдается сильное смещение в сторону делеций; при отсутствии отбора происходило бы сокращение генома на 0.3 нт на каждую однонуклеотидную замену.
- 2) Действие отрицательного и положительного отбора препятствует сокращению генома. Среди новых мутаций доля отсеиваемых отрицательным отбором

составляет 71% для инсерций и 82% для делеций. Среди мутаций, закрепляющихся в межвидовой эволюции, доля закреплённых под действием положительного отбора составляет 67% для инсерций и 36% для делеций.

3) Инсерции и делеции подвергаются действию генной конверсии, смещённой в сторону инсерций. Смещение вероятности закрепления соответствует усреднённому по геному конверсионному преимуществу для инсерций $N_e\omega = \sim 0,3-0,4$.

4) Скорость эволюции в синонимических сайтах не меняется после возникновения индела. Из этого следует, что инделы не обладают мутагенным эффектом на окружающую последовательность.

5) Скорость эволюции в несинонимических сайтах значительно возрастает после возникновения индела, причём этот эффект сильнее для делеций. Несинонимичные замены, вызванные инделами, носят исключительно адаптивный характер.

6) Событие инсерции вызывает приблизительно 1 дополнительную аминокислотную замену, а событие делеции – приблизительно 5 дополнительных аминокислотных замен. Различие между инсерциями и делециями, вероятно, связано с тем, что делеции снижают приспособленность сильнее инсерций.

7) Анализ инделов разных возрастов показывает, что адаптивная прогулка происходит за время, которое соответствует 0,1–0,7 синонимическим заменам.

8) Среди замен, которые произошли после события индела, доля тех, что произошла в консервативных сайтах, была выше, чем для замен, не связанных с инделами. Для делеций это различие выражено сильнее, чем для инсерций.

Список публикаций по теме диссертации

Статьи в изданиях, рекомендованных перечнем ВАК

Leushkin EV, Bazykin GA, Kondrashov AS. Insertions and deletions trigger adaptive walks in *Drosophila* proteins // *Proc Biol Sci.*- 2012. - Vol. 279. - P. 3075-3082

Leushkin EV, Bazykin GA, Kondrashov AS. Strong mutational bias towards deletions in the *Drosophila melanogaster* genome is compensated by selection // *Genome Biol Evol.* - 2013. - Vol. 5. - P. 514-524

Leushkin EV, Bazykin GA. Short indels are subject to insertion-biased gene conversion // *Evolution.* - 2013. - Vol. 67. - P. 2604-4613

Труды конференций

Leushkin EV, Bazykin GA, Kondrashov AS. Adaptive amino acid replacements triggered by indels in *Drosophila* proteins // MCCMB 2011: Proceedings of the International Moscow Conference on Computational Molecular Biology. - Moscow, Russia, 2011. - P. 198

Leushkin E.V., Bazykin G.A., Kondrashov A.S. Insertions and deletions trigger adaptive walks in *Drosophila* proteins. // Otto Warburg International Summer School and Research Symposium 2011 on Evolutionary Genomics. - Berlin, Germany, 2011.

Leushkin EV, Kondrashov AS, Bazykin GA. Insertion-biased gene conversion for short indels // SMBE 2012: Annual Meeting of the Society for Molecular Biology and Evolution. - Dublin, Ireland, 2012. - P. 1082

Leushkin E.V. Bazykin G.A.. Insertion-biased gene conversion for short indels. // ITaS 2012: Information Technology and Systems – 2012. - Petrozavodsk, Russia, 2012. - P. 319