Федеральное государственное бюджетное учреждение науки Институт проблем передачи информации им. А.А. Харкевича Российской академии наук

На правах рукописи

Тереханова Надежда Владимировна

# Неравномерность мутагенеза и отбора в геноме позвоночных

03.01.09 - математическая биология, биоинформатика

Диссертация на соискание ученой степени кандидата биологических наук

Научный руководитель:

к. б. н. Базыкин Г. А.

## Оглавление

Введение	4
Актуальность работы	4
Цели и задачи исследования	5
Научная новизна и практическая значимость	5
Положения, выносимые на защиту	6
Степень достоверности и апробация результатов	7
Структура и объём диссертации	7
Обзор литературы	8
1. Типы мутаций и методы измерения локальной скорости мутирования (ЛСМ)	8
2. Факторы, влияющие на ЛСМ	9
3. Геномная архитектура адаптации	18
Глава 1. Эволюция локальной скорости мутирования (ЛСМ) и определяющих её факторов	в 26
1.1. Материалы и методы	26
1.1.1. Выравнивания	26
1.1.2. Картирование геномных характеристик	27
1.1.3. Вычисление ЛСМ	28
1.1.4. Объясняемая дисперсия ЛСМ	29
1.1.5. Корреляция изменения ЛСМ с изменениями геномных свойств	29
1.1.6. Поиск участков генома с ускоренной или замедленной ЛСМ	29
1.2. Результаты и обсуждение	30
1.2.1. ЛСМ сильно скоррелированы между человеком и другими приматами	30
1.2.2. Объяснение дисперсии ЛСМ при помощи геномных свойств	42
1.2.3. Изменения скорости рекомбинации связаны с изменениями ЛСМ	48
1.3. Выводы	58
Глава 2. Геномная архитектура адаптации у трёхиглой колюшки	60
2.1. Материалы и методы	60

2.1.1. Сбор образцов и геномное секвенирование 60	0
2.1.2. Геномное картирование и анализ последовательностей	2
2.1.3. Кластеры маркерных однонуклеотидных полиморфизмов в геноме	2
2.1.4. Валидация частот пресноводных аллелей в островах дивергенции (ОД) 63	3
2.1.5. Оценка коэффициентов отбора в ОД 63	3
2.2. Результаты и обсуждение 64	4
2.2.1. Генетические различия между морской и пресноводными популяциями	4
2.2.2. Динамика адаптации G. aculeatus к пресноводной среде обитания	2
2.2.3. Сила отбора, способствующего адаптации к пресноводной среде обитания	0
2.3. Выводы	3
Список публикаций по теме диссертации	4
Тезисы конференций	4
Благодарности	5
Список сокращений	5
Список литературы	6

## Введение

## Актуальность работы

Мутагенез и естественный отбор являются основными факторами эволюции. С развитием технологий секвенирования появилось большое количество геномных данных для разных видов и некоторых популяций, что открывает новые возможности для изучения мутагенеза и отбора на уровне отдельных сегментов генома.

Нуклеотидная последовательность ДНК изменяется в ряду поколений в результате мутагенеза. Мутации бывают разных типов и могут быть вызваны действием различных факторов. Поэтому важно понимать вклады различных процессов, которые участвуют в возникновении мутаций, а также учитывать неравномерность действия этих факторов в разных сегментах последовательности ДНК. В результате действия отбора мутации могут быть в разной степени вредными или полезными, или же они могут быть нейтральными. Данное свойство влияет на динамику частоты мутации внутри популяции, и видимые изменения ДНК являются результатом совместного действия мутагенеза и отбора.

Естественный отбор является процессом, который отсеивает изменения ДНК, ухудшающие приспособленность организма, и благоприятствует закреплению в популяции полезных вариантов, улучшающих приспособленность. При изменении условий окружающей среды меняется действие отбора, что может привести к тому, что определённые мутации могут стать более полезными в новой среде. Такие мутации могут либо уже присутствовать в популяции, либо возникать *de novo*. Понятно также, что мутации, обеспечивающие адаптацию, будут оказывать влияние на гены, изменение функций которых является важным при переходе в новую среду. На сегодняшний день известно, что в результате адаптации популяции происходит накопление различий между геномами популяций [1,2]. Было показано, что в результате действия дизруптивного отбора в геномах образуются участки с повышенным количеством различий между популяциями ещё до момента полной репродуктивной изоляции видов [3].

Биоинформатический анализ нуклеотидных последовательностей молекул, полученных в результате секвенирования, позволяет делать выводы о закономерностях их структуры и эволюции. На сегодняшний день известно, что геном имеет организованную многоуровневую структуру, обеспечивающую компактное хранение большого количества ДНК в ядре. Накопление мутаций связано с повреждением ДНК, а также с молекулярными процессами репликации, репарации, рекомбинации и транскрипции, при протекании которых возникают ошибки. Факторы, вовлечённые в эти процессы, имеют свои особенности, что приводит к неравномерному накоплению мутаций в разных сегментах генома. В результате большого количества исследований, в том числе таких больших проектов, как ENCODE [4] и RoadMap [5], построено множество карт, описывающих расположение генов, структурных элементов хроматина и другие свойства сегментов ДНК.

На теоретическом уровне понимание различий скоростей мутирования и отбора между сегментами генома необходимо для построения более точных эволюционных моделей и лучшего понимания процесса видообразования, в том числе для нашего собственного вида. Помимо теоретической значимости, изучение неравномерности мутагенеза между сегментами генома имеет прикладное значение, например, в медицинской генетике. Поиск генов, ответственных за генетические заболевания, и генов-драйверов при мутагенезе требует построения правильной нуль-модели, учитывающей неравномерность скорости накопления соответственно мутаций в зародышевой линии и соматических мутаций.

## Цели и задачи исследования

Целью исследования было изучение закономерностей различий скоростей мутирования и действия отбора в разных сегментах генома, а также описание факторов, которые оказывают влияние на эти процессы. Были поставлены следующие задачи:

1. Изучить закономерности изменения локальной скорости мутирования в ходе эволюции нескольких видов приматов и грызунов, а также определить причины этих изменений.

2. Исследовать структуру и свойства участков повышенной межпопуляционной изменчивости («островов дивергенции»), которые возникают вследствие действия отбора, на примере нескольких независимых популяций трёхиглой колюшки, и оценить силу отбора, приведшего к их возникновению.

#### Научная новизна и практическая значимость

В предыдущих работах было показано, что локальная скорость мутирования (ЛСМ) является схожей в гомологичных сегментах геномов близкородственных видов, но сильно отличается у давно разошедшихся видов [6,7]. Эти результаты были получены на малом числе видов, вследствие чего было трудно разобраться, как быстро распадается корреляция ЛСМ между близкородственными видами с ростом филогенетического расстояния. В настоящей работе были использованы данные из выравнивания 100 видов позвоночных [8], которое появилось в открытом доступе в 2014 году. Был проведён анализ ЛСМ при использовании филогении 9 видов приматов, а также были детально проанализированы вклады различных факторов мутагенеза в

изменение ЛСМ на разных эволюционных расстояниях. Были качественно и количественно оценены изменения ЛСМ между близкородственными видами приматов, найдены закономерности этих изменений и оценены вклады в изменение ЛСМ различных факторов мутагенеза. Полученные результаты в дальнейшем могут быть использованы при разработке эволюционных моделей.

Изучение неравномерности отбора также является важной задачей для понимания процессов, связанных с изменениями в геноме. В 2005 году было показано существование геномных участков с повышенным количеством различий между адаптирующимися к разным условиям популяциями (островов дивергенции, ОД, [3]). С введением в широкую практику методов высокопроизводительного секвенирования было показано наличие ОД во множестве других пар популяций, адаптирующихся к разным условиям среды. Тем не менее, на сегодняшний день факторы, влияющие на возникновение и структуру ОД, а также их роль в процессе видообразования остаются малопонятными. Существует круг модельных организмов для данных типов исследований, и к ним относится трёхиглая колюшка *Gasterosteus aculeatus*. В данной работе впервые разработан метод поиска ОД при использовании данных полногеномного секвенирования объединенных выборок (pool-seq) популяций колюшки. Оценены изменения частоты пресноводного аллеля в ОД в молодых пресноводных популяциях, в том числе созданных экспериментально. Количественно оценена сила отбора, действующего на ОД.

#### Положения, выносимые на защиту

1. Показано, что учёт локальной скорости мутирования (ЛСМ) близкородственных видов приматов позволяет объяснить существенно (почти в два раза) большую долю дисперсии ЛСМ человека по сравнению с использованием только карт известных геномных разметок, как это делалось в предыдущих работах.

2. Вклад скорости рекомбинации в объясняемую дисперсию ЛСМ является наибольшим среди всех исследованных геномных характеристик. Однако, если для предсказания ЛСМ в одном виде используется локальная скорость рекомбинации другого вида, то её предсказательная сила быстро падает с ростом филогенетического расстояния между видами. Изменения скорости рекомбинации между видами человек и мышь связаны с изменениями ЛСМ сильнее, чем изменения других исследованных геномных характеристик.

3. На примере человека и шимпанзе разработан метод поиска участков генома с повышенной и пониженной ЛСМ на линии вида. Показано, что участки с повышенной ЛСМ у человека или у шимпанзе в среднем имеют повышенные значения скоростей рекомбинации.

Данный результат свидетельствует о том, что рекомбинация ассоциирована с изменением мутагенеза на коротких эволюционных временах.

4. Описаны участки генома, которые предположительно вовлечены в адаптацию к пресноводной среде обитания – острова дивергенции (ОД). Найдены 19 ОД, содержащих повышенное количество различий между морскими и пресноводными особями трёхиглой колюшки. Показано, что скорость эволюции повышена в ОД, и изучена динамика изменения частот пресноводных аллелей внутри ОД. Полученные результаты согласуются с гипотезой о том, что для адаптации используется предсуществующая генетическая изменчивость.

5. Показано, что на ОД при адаптации действует сильный отбор. Средние значения коэффициентов отбора составили *s*=0,16 для популяции озера Ершовское и *s*=0,13 для популяции карьера Голубой.

## Степень достоверности и апробация результатов

По материалам диссертации опубликовано 2 статьи в рецензируемых научных журналах. Результаты работы были представлены на международных конференциях SMBE'12, MCCMB'13, SMBE'14, MCCMB'15 и российских конференциях ИТИС'12, ИТИС'13, ИТИС'15.

#### Структура и объём диссертации

Диссертация состоит из введения, обзора литературы, 2 глав и библиографии. Общий объем диссертации 94 страницы, из них 85 страниц текста, включая 45 рисунков и 8 таблиц. Библиография включает 131 наименование на 9 страницах.

## Обзор литературы

Внутривидовая и межвидовая изменчивость генома представляют собой результат длительной эволюции под действием факторов, включающих мутагенез и отбор. Характер и интенсивность этих процессов различаются между сегментами генома [3,6], что влияет на плотность наблюдаемых мутаций и их частоты. Определение характеристик этих процессов представляется важной и сложной задачей. При этом дополнительную информацию о причинах неравномерности накопления и распространения вариантов можно получить, используя дополнительные знания о структуре генома (карты участков гетерохроматина, горячих точек рекомбинации, расположения генов и другие).

В главе 1 исследуется изменение с ходом эволюции локальной скорости мутирования (ЛСМ), а также оценивается влияние геномных факторов на ЛСМ и её эволюцию. Один из примеров прикладного применения знаний о неравномерности ЛСМ – поиск драйверов раковых заболеваний [9]. В ранних исследованиях в качестве драйверных генов ошибочно выделялись те, что находились в быстро мутирующих участках, такие как гены рецепторов обоняния. Учёт изменчивости ЛСМ позволил значительно улучшить качество определения драйверов.

В течение всего жизненного цикла генетического варианта, например – однонуклеотидного полиморфизма (ОНП), от момента его возникновения в результате мутации до исчезновения или закрепления, на него может действовать естественный отбор. Различия в интенсивности действия отбора еще сильнее варьируют между сегментами генома, чем различия в интенсивности мутагенеза, поскольку сила и направление отбора полностью определяются функцией соответствующего участка. Одной из наиболее интересных форм отбора является тот, который можно наблюдать при адаптации к новым условиям среды. В главе 2 исследуются участки с повышенным количеством различий (острова дивергенции, ОД) между морскими и пресноводными популяциями трёхиглой колюшки, в которых предположительно находятся мишени для действия отбора. Были идентифицированы ОД для популяций Белого моря, описаны их свойства, получена количественная оценка силы отбора, действующего на них. Совместное изучение неравномерности мутагенеза и геномной архитектуры адаптации позволяет понять взаимодействие факторов, влияющих на эволюцию генома и определяющих его изменчивость.

## 1. Типы мутаций и методы измерения локальной скорости мутирования (ЛСМ)

Накопление мутаций обусловлено как внешними, так и внутренними для организма факторами. Внешние факторы включают воздействие мутагенов, таких как радиация, химические реагенты и другие. К последствиям внутренних факторов относятся мутации, возникающие в результате ошибок важных молекулярных процессов, таких как репликация, транскрипция, репарация и рекомбинация. Различные молекулярные процессы характеризуются различными вызываемыми ими паттернами мутаций. Существуют методы, позволяющие раскладывать наборы мутаций, наблюдаемых в геноме, на вызывающие их мутационные процессы, выделяя вклады отдельных молекулярных процессов в их возникновение [10].

Мутации могут быть разделены на два типа: соматические и зародышевой линии. Первый тип составляют мутации, которые накапливаются в клетках организма в течение жизни и не передаются по наследству. Ко второму типу относятся мутации, которые происходят в клетках зародышевой линии и являются наследуемыми. Мутации обоих типов могут вызывать заболевания. Мутации первого типа характерны для раковых заболеваний, проявляющихся, как правило, с возрастом. Мутации второго типа вызывают наследственные заболевания. Следует отметить, что эти два типа мутаций обладают разными характеристиками и изучаются на разных данных. В данной работе мы будем рассматривать мутации наследственные.

При изучении мутаций зародышевой линии используется несколько подходов с соответствующими типами данных: данные по дивергенции видов, данные по внутривидовому полиморфизму и данные по тройкам семей (мать, отец и ребёнок). Наиболее прямым и точным методом является метод с использованием троек семей, когда мутации непосредственно детектируются в потомке по отношению к его родителям. Тем не менее, у этого метода есть свои недостатки, наиболее критичным из которых является малое количество мутаций, детектируемых в одной семье, что затрудняет получение высокой статистической значимости при исследовании ЛСМ таким способом. При использовании двух других методов доступно гораздо больше событий мутаций, так как рассматриваются большие времена. Недостатками методов с использованием данных по дивергенции и по полиморфизму является то, что закрепление наблюдаемых мутаций происходило с участием действия отбора и других факторов. Поэтому желательно при использовании этих подходов учитывать их возможный вклад. Непосредственно в данной работе использовались данные последовательностей близких видов приматов с добавлением данных по полиморфизму в человеческой популяции. Также полученные результаты проверялись на данных мутаций, детектированных по методу с использованием троек семей.

#### 2. Факторы, влияющие на ЛСМ

Скорость мутирования неравномерна вдоль генома (рисунок 1 и [11,12]). Причиной неравномерности является неоднородность многочисленных молекулярных процессов, которые происходят с участием ДНК, в том числе репликации, транскрипции, репарации и рекомбинации.

9

Данные процессы связаны с работой специальных белковых факторов, которые имеют сродство к специфическим последовательностям – сайтам связывания. Тем самым, распределение таких сайтов влияет на распределение данных процессов и их динамику. Также на ЛСМ влияет строение хроматина, включающее такие свойства, как распределение нуклеосом, модификации гистонов и степень открытости участков ДНК [13].



Рисунок 1. Изменчивость ЛСМ человека вдоль хромосомы 8 (по [11] с изменениями). По вертикальной оси - дивергенция человека и шимпанзе. Более тёмная штрихованная линия обозначает среднее значение по геному, более светлые штрихованные линии соответствуют двум стандартным отклонениям от среднего значения.



Рисунок 2. Факторы, связанные с устройством хроматина, влияющие на ЛСМ. Закрытый хроматин может быть «молчащим», с практически не меняющимся распределением гистоновых меток, или же репрессированным с модификациями H3K9me3 или H3K27me3. ДНК, которая активно регулируется и транскрибируется, находится в участках открытого хроматина, чувствительных к ДНКазе. Такие участки имеют повышенную плотность сайтов посадки факторов транскрипции, а также гистоновых модификаций: H3K4me1 (связанной с энхансерами), H3K4me3 (связанной с промоторами), H3K36me3 (связанной с транскрибируемым хроматином), H3K27ac и H4ac (связанными с энхансерами и промоторами) (по [13] с изменениями).

Важным процессом клетки, который влияет на ЛСМ, является репликация ДНК. У эукариот данный процесс начинается сразу в нескольких точках, ориджинах, с образованием двусторонней вилки репликации, движущейся в двух противоположных направлениях вплоть до момента слияния со встречными вилками. Таким образом, некоторые участки генома реплицируются раньше по сравнению с другими участками, и было показано, что в таких участках в среднем накапливается меньшее количество мутаций (рисунок 3, [14]). Данное наблюдение было сделано для различных классов мутаций, что говорит о наличии общего механизма. В качестве возможной причины накопления мутаций в более поздно реплицирующихся участках является их более долгое нахождение в одноцепочечном состоянии, так как ДНК в таком состоянии более подвержена эндогенному разрушению.



Рисунок 3. Зависимость плотности ОНП от времени репликации участков, в которых они были измерены (по [14] с изменениями).

Другим важным процессом, происходящим в клетке, является рекомбинация, которая происходит при образовании гамет в процессе мейоза. Рекомбинация связана с возникновением двуцепочечных разрывов в ДНК и образованием структуры Холлидея. Затем следует разрезание данной структуры, несколькими способами, и только в некоторых случаях оно приводит к образованию кроссоверов (рисунок 4, [15]). Было показано, что процесс рекомбинации связан с повышенным накоплением мутаций [16]. Также известно, что распределение горячих точек рекомбинации достаточно быстро меняется с течением времени. Данный эффект наблюдается не только между близкими видами, как например человек и шимпанзе (коэффициент корреляции скоростей рекомбинации между этими двумя видами составляет ~50% [17]), но также заметен даже среди индивидуумов одного вида, как например у шимпанзе (рисунок 5, [17]) и человека

(рисунок 6, [18]). Изменение скорости рекомбинации может быть связано с быстрой эволюцией *PRDM9*, что было, например, показано для человека [18].



Рисунок 4. Кроссинговер хромосом и генная конверсия в процессе мейоза [15].



Рисунок 5. Соотношение корреляции скорости рекомбинации и дивергенции между популяциями и видами некоторых приматов. По горизонтальной оси указаны попарные значения дивергенции между видами или популяциями. По вертикальной оси указан коэффициент корреляции Спирмена между значениями рекомбинации внутри 1 Мб геномных окон для тех же сравнений. Афр. (YRI) – африканская популяция человека, Евр. (CEU) – европейская популяция человека (по [17] с изменениями).



Рисунок 6. Расположение горячих точек рекомбинации может различаться у индивидуумов популяции человека [18]. Показаны распределения двуцепочечных разрывов индивидуумов, которые имеют различные аллели гена *PRDM9*, внутри 300 Кб участка хромосомы 17. На панелях вертикальная ось соответствует значениям покрытий однонитевой ДНК, которая использовалась в качестве маркера разрывов, нормализованных на количество фрагментов на Кб участка, на миллион геномных чтений (FPKM).

Помимо своей мутагенности, процесс рекомбинации связан с генной конверсией. Было показано, что только ~10% двуцепочечных разрывов завершаются образованием кроссоверов у млекопитающих [19]. При генной конверсии происходит замещение небольшой части нуклеотидов одной из цепей молекулы ДНК на нуклеотиды гомологичной последовательности. В некоторых случаях в результате данного процесса будет возникать ситуация, когда два нуклеотида молекулы ДНК не будут комплементарны, и тогда на место несоответствия будут привлечены белки системы репарации. Было установлено, что в такой ситуации в большем числе случаев будет отдаваться предпочтение заменам А/Т→Г/Ц по сравнению с противоположными заменами [16]. Данный эффект носит название ГЦ-смещённой генной конверсии.

Транскрипция также является важным процессом клетки, влияющим на скорость накопления мутаций. Так, известно, что РНК-полимераза способна «застревать» на участках неспаренных нуклеотидов, призывая при этом комплекс белков к участку с мутацией, которые осуществляют репарацию ДНК. Таким образом, участки с повышенным содержанием активно транскрибирующихся генов имеют дополнительную систему исправления ошибок за счёт этого механизма (рисунок 7, [20]). На данных образцов раковых тканей также было показано, что имеет место повреждение ДНК, ассоциированное с транскрипцией, которое происходит на не транскрибируемой цепи [21].



Рисунок 7. Модель репарации ДНК, ассоциированной с транскрипцией. На рисунке показаны РНК-полимераза II, которая во время транскрипции движется вдоль последовательности ДНК, взаимодействия с белками CSB и UVSSA/USP7. Полимераза может «застрять» в участке с неспаренными нуклеотидами ДНК, что способствует стабилизации связи CSB и полимеразы. В результате этого происходит сборка комплекса UVSSA/USP7 в месте повреждения (по [20] с изменениями).

При отсутствии активности ДНК находится в структурированном состоянии, связанной с белками гистонами с образованием организованной структуры хроматина. Всего существует 5 типов гистонов, которые в комплексе связываются с ДНК и образуют нуклеосомы. Данные белки очень консервативны, что подчёркивает их исключительное значение в поддержании целостности и сохранности генома. Различные участки генома имеют несколько различающиеся плотности нуклеосом, что влияет на ЛСМ. В частности, было показано, что повышенное содержание нуклеосом ассоциировано с менее эффективной репарацией [22]. Кроме того, у нуклеосом есть выступающие за их пределы достаточно длинные N-концевые фрагменты гистоновых белков (рисунок 8). Модификации данных хвостов играют важную роль в структурировании хроматина и регуляции экспрессии генов. Было показано, что модификации определённых остатков гистонов несут специфические функции. Главными модификациями являются метилирование и ацетилирование. На сегодняшний день существуют карты различных модификаций гистоновых белков (рисунок 2).



Рисунок 8. Строение хроматина и нуклеосомы (по [15]).

Более высоким уровнем организации хроматина является наличие топологически ассоциированных доменов, которые представляют собой 3D-структуры линейных участков ДНК [23]. Для таких структур количество внутренних контактов превышает количество внешних. Одним из наиболее современных методов определения топологически ассоциированных доменов является анализ Hi-C, который позволяет выявить все контакты между различными участками генома. Было показано, что топология таких структур также может влиять на ЛСМ [24].

Несколько других свойств генома, таких как распределение генов, консервативность и др., связаны с действием отбора [25]. Известно, что гены и регуляторные последовательности находятся под действием отбора более сильным по сравнению с большинством некодирующих последовательностей, и их распределение также влияет на оценку ЛСМ при использовании данных полиморфизма или дивергенции.

Стоит отметить, что множество описанных процессов и факторов не являются независимыми. Напротив, карты их распределений коррелируют с высоким уровнем значимости. Так, например, участки ранней репликации обогащены генами и имеют повышенное содержание Г и Ц нуклеотидов (ГЦ-состав) [26]. Также известно, что рекомбинация происходит в участках с повышенным ГЦ-составом [27]. Таким образом, молекулярные процессы взаимосвязаны друг с другом, и одни и те же факторы могут оказывать воздействие на множество процессов и влиять друг на друга. Вследствие этого, важным моментом при выявлении причин мутагенеза является оценка независимых вкладов геномных свойств в объяснение мутагенеза.

На данный момент существуют карты геномных свойств, разрешение которых достигает тысяч и сотен нуклеотидов. Данные разметки описывают свойства структуры генома, и взаимосвязи с протекающими молекулярными процессами. Изучался вклад в изменение ЛСМ таких свойств генома, как время репликации [26]; плотность сайтов, чувствительных к ДНКазе [28]; плотность генов; ГЦ-состав; различные модификации гистонов [6,29] и другие. Тем не менее, даже если учесть все доступные разметки, то они будут определять менее 30% изменчивости ЛСМ в геноме человека (рисунок 9, [6]). Исходя из этого, можно предположить, что пока описаны ещё не все разметки, вносящие вклад в объяснение ЛСМ.



Рисунок 9. Предсказание ЛСМ при использовании различных геномных разметок методом линейной регрессии. Доля объясняемой дисперсии измеряется величиной R<sup>2</sup>. По вертикальной оси показан кумулятивный R<sup>2</sup>, полученный при регрессии с использованием факторов, указанных по горизонтальной оси, включая все те, что расположены слева (по [6] с изменениями).

Из других работ известно, что ЛСМ является схожей у близкородственных видов [7], но практически совершенно разной у эволюционно более далёких видов [30]. По-видимому, это означает, что карты разметок геномных свойств являются более схожими у близкородственных

видов по сравнению с менее родственными. Тем не менее, детальное изучение изменений карт разметок в процессе эволюции не проводилось. Динамика изменчивости ЛСМ и причины её изменений до сих пор плохо понятны [7,12], и количество объясняемой дисперсии ЛСМ не слишком велико. Также, вероятно, могут существовать какие-либо ещё не открытые свойства ДНК, которые могут влиять на ЛСМ.

Эволюция и влияние друг на друга ЛСМ и геномных свойств могут быть изучены с использованием метода корреляций между ЛСМ и геномными свойствами в гомологичных окнах между видами с известной филогенией. Для такой задачи желательно использовать виды, не слишком давно разошедшиеся друг от друга после ответвления от общего предка, чтобы уменьшить вклад побочных факторов, таких как, например, возвратные мутации. С другой стороны, для очень близких видов стоит учитывать неполную сортировку ветвей, проявляющуюся в том, что для отдельных участков генома филогения будет не соответствовать канонической, полученной по наибольшему числу участков генома, что может привести к ошибкам при детектировании мутаций. Кроме того, общий полиморфизм, характерный для очень близких видов, может приводить к завышению оценки корреляции ЛСМ по данным полиморфизма. Для близкородственных видов человека и шимпанзе неполная сортировка ветвей характерна для ~30% участков генома [31].

Оценки мутагенеза по данным полиморфизма и дивергенции подвержены смещению вследствие действия отбора. Для проверки результатов в нашей работе использовались данные по *de novo* мутациям из троек семей. Для контроля также использовались разметки, которые скоррелированы с действием отбора, такие как частота минорного аллеля, плотность экзонных нуклеотидов, консервативность и другие. Распределение функциональных элементов неравномерно в геноме, из чего следует, что действие отбора также имеет неравномерный характер.

#### 3. Геномная архитектура адаптации

Неравномерность действия отбора вдоль генома, проявляющаяся при адаптации популяции к новым условиям окружающей среды, играет особую роль в видообразовании. Видообразование является сложным процессом, который до сих пор не до конца понятен (рисунок 10). Оно может происходить как с наличием географической изоляции (аллопатрическое), так и при её отсутствии (парапатрическое или симпатрическое). В обоих случаях две возникающие популяции начинают накапливать адаптивные различия вследствие приспособления к разным условиям. Таким образом, отбор будет благоприятствовать возникновению и закреплению в популяции фенотипических различий между популяциями, адаптированными к разным

условиям. При большой степени дивергенции генома один вид может разделиться на два, индивидуумы которых уже не смогут образовывать плодовитое потомство. Кроме того, отбор будет способствовать возникновению препятствий для образования гибридного потомства – репродуктивной изоляции. У позвоночных распространена предзиготическая изоляция, когда индивидуумы способны отличать представителей своего вида от представителей другого по фенотипическим признакам и отдают предпочтение при скрещивании особям своей популяции. Такая способность даёт селективное преимущество, потому что гибриды, несмотря на то что в части случаев могут быть жизнеспособны, тем не менее, являются менее приспособленными к той или иной среде обитания, чем их родители. Открытым вопросом остаются точные причины возникновения репродуктивной изоляции. Тем не менее, было показано, что в среднем при ~1% дивергенции между популяциями их индивидуумы теряют способность образовывать плодовитое потомство при межпопуляционном скрещивании (рисунок 10).

Важная стадия процесса видообразования происходит в момент, когда популяции приобрели адаптацию к разным условиям, но всё ещё являются одним видом. Существует множество примеров этой стадии [32–34], в том числе и у человека [35]. Адаптировавшиеся к разным условиям популяции называют подвидами, морфами, расами. Изучение процесса видообразования на ранних стадиях является важным для понимания эволюционных процессов и их влияния на изменение генома. Исследования такого рода имеют долгую историю [36], и сейчас с доступными методами высокопроизводительного секвенирования нового поколения появляется возможность улучшить понимание данного процесса.



Расстояние по синонимическим заменам

Рисунок 10. Накопление различий между видами или популяциями при различных вероятностях миграций. Каждая точка соответствует одной паре видов/популяций животных [37]. Горизонтальная ось соответствует среднему генетическому расстоянию, измеренному в синонимических заменах (логарифмическая шкала). Вертикальная ось соответствует вычисленной вероятности миграций между рассматриваемыми видами/популяциями пары (величиной обратно пропорциональной степени репродуктивной изоляции). Для каждой пары рассматривалось несколько моделей, которые описывали наблюдаемые данные для популяций/видов. Цвет точек соответствует результатам, полученным из расчётов: красный – пары с высокой вероятностью изоляции между ними, серый – пары с неизвестным статусом изоляции, фиолетовый – пары со свидетельствами неполной изоляции, голубой – пары с высокой вероятностью отсутствия изоляции (по [37] с изменениями).

Представители подвидов адаптируются к разным условиям среды, и поэтому логично предположить, что они будут иметь различия в геномах. Также понятно, что при скрещивании между особями, адаптированными к разным условиям, количество различий между популяциями будет снижаться. Накопившиеся геномные перестройки в каждой отдельной популяции будут локально препятствовать этому процессу. При изучении подвидов было неоднократно показано, что существуют участки, которые имеют повышенное содержание отличий между представителями двух подвидов, которые получили название островов дивергенции (ОД, рисунок 11) [32–34]. Объяснением природы таких участков может являться наличие в них генов или регуляторных областей, которые обеспечивают адаптацию к конкретной среде обитания. Следовательно, данные участки находятся под дивергентным (дизруптивным) отбором, который благоприятствует одному варианту в одной среде, а другому – в другой.



Рисунок 11. Неравномерность накопления различий между близкородственными видами *H. m. aglaope* и *H. m. amaryllis*, острова дивергенции. В качестве примеров выбраны участки, ответственные за окрашивание крыльев (HmYb участок (a), HmB/D участок (b)). По горизонтальной оси указаны координаты вдоль хромосом, по вертикальной оси указана мера генетического различия, измеренная при помощи F<sub>ST</sub>-статистики (по [38]).

Механизмы образования ОД не до конца понятны. Существует теория «дивергенции с хитчхайкингом» (divergence hitchhiking), согласно которой при наличии неполной изоляции между адаптирующимися к разным условиям популяциями новая полезная мутация будет иметь больший шанс на закрепление в популяции при условии, если она возникнет поблизости от другой полезной мутации [1]. Тем самым, возможно последовательное накопление таких мутаций, что будет приводить к образованию и последующему удлинению ОД с течением времени (рисунок 12). Близость полезной мутации сильного эффекта снижает вероятность для слабо полезной мутации быть потерянной в результате дрейфа во время фиксации. С другой стороны, при отсутствии изоляции популяций, часть подобных связей между мутациями будет разобщаться, а границы ОД будут размываться. До конца непонятно, насколько данная гипотеза образования ОД верна. Существует работа, в которой показано, что одной только близости по геному полезных мутаций недостаточно для образования ОД при большинстве эволюционных сценариев [39]. Другим аргументом против этой теории является наблюдаемое снижение рекомбинации в участках ОД и наличие инверсий в составе некоторых ОД. На основе этих результатов выдвигаются предположения, что в участках ОД присутствует некоторое количество структурных перестроек, которые будут затруднять или прекращать полностью разобщение

мутаций [2,39]; в этих случаях механизм «дивергенции с хитч-хайкингом» не нужен. Также есть работа, показывающая при помощи вычислительных экспериментов, что при дивергентном отборе участки с большими коэффициентами отбора, будут более устойчивы к дрейфу и иметь меньшую вероятность быть потерянными в процессе фиксации [40]. Тем не менее, вопрос образования ОД и их свойства являются до сих пор недостаточно изученными.



Рисунок 12. Теория «дивергенции с хитч-хайкингом» образования и роста ОД. Различия в ДНК между двумя популяциями, которые предоставляют селективное преимущество, обозначены зелёным цветом (по [41] с изменениями).

Экспериментальное изучение адаптации у позвоночных затруднено большой длиной поколения, замедляющей процессы эволюции. Тем не менее, был осуществлен ряд эволюционных экспериментов по адаптации у позвоночных [42–44]. Поскольку большинство экспериментов относительно кратковременны, действие отбора должно быть достаточно сильным, чтобы его эффект был заметен. Экспериментальная эволюция с последующим секвенированием применяется широко при изучении бактерий вследствие того, что они имеют короткий цикл воспроизведения. К тому же небольшой геном дополнительно облегчает изучение эволюции для данных объектов [45,46]. Подобные исследования проводятся на позвоночных, но гораздо реже [42,47]. Также результаты действия положительного отбора могут быть видны не только на паттернах полиморфизма при сравнении генотипов одного вида, но и на уровне дивергенции видов. Механизмы адаптации на генетическом уровне всё ещё плохо изучены [48,49].

Исследования геномики адаптации получили сильное развитие с введением в широкую практику методов высокопроизводительного секвенирования нового поколения. Стала возможной идентификация участков под отбором и детальное изучение паттернов дивергенции

и полиморфизма и в таких участках, и вокруг них, в ряде модельных объектов [43,50,51]. В исследованиях такого рода ведётся поиск участков под отбором, а также рассчитывается количество замен между популяциями в генах и регуляторных областях. Бывает, что одна и та же адаптация возникает несколько раз в нескольких независимых популяциях, но частота таких параллельных изменений сильно варьирует в разных системах [32,43,52]. До сих пор нет объяснений причин этих различий, поэтому необходимы дальнейшие исследования в этом направлении.

Трёхиглая колюшка (*Gasterosteus aculeatus*) является широко используемым модельным организмом при изучении адаптации и видообразования [32,53,54]. Данный вид имеет очень большое количество разнообразных морф [55,56]. В данное время, исходя из результатов ряда анализов, считается, что предковая популяция колюшки жила в морской воде и что колонизация пресноводных водоёмов с последующей адаптацией к пресной воде в них происходила множество раз по всему Северному полушарию [57]. Индивидуумы морской популяции используют пресноводные озёра только в качестве временных мест пребывания для постройки гнёзд при размножении. В то же время были основаны тысячи независимых озёрных популяций, и индивидуумы в них значительно дивергировали на разных уровнях, включая морфологию, физиологию и поведение, приобретая способность выживать в новой среде в течение всей жизни. Независимое возникновение пресноводных популяций *G. aculeatus* в различных водоёмах Северного полушария даёт возможность изучать параллельные случаи возникновения адаптации при схожих условиях среды обитания [32,58,59].

Было показано, что большая часть генетических изменений при адаптации является параллельной за счёт повторяемого использования одних и тех же аллелей из предсуществующей генетической изменчивости (рисунок 13, [52]). Было также выяснено, что возраст ОД является достаточно большим и в разы превышает возрасты пресноводных популяций, что свидетельствует о том, что в адаптации участвует древний полиморфизм [60]. Аллели адаптации к пресной воде находятся на низкой частоте в морской популяции [54]. Вероятно, каждая отдельная особь из набора индивидуумов, которые являются основателями новой пресноводной популяции, несёт варианты, полезные в пресной воде, только в малой части участков ОД, так как эти варианты неблагоприятны в морской среде. Тем не менее, попав в пресноводную среду, особи с такими вариантами аллелей будут иметь преимущество при отборе, и вследствие этого частота этих вариантов внутри популяции будет быстро расти. За счёт этого колюшка может очень быстро адаптироваться к пресной воде, и пресноводные аллели постоянно присутствуют в морской популяции на низких частотах [61], предположительно в результате миграций из пресноводных популяций. Вследствие этого ожидается, что эти аллели, обеспечивающие

адаптацию к пресноводной среде обитания, будут одинаковыми в независимо образовавшихся пресноводных популяциях, а морские и пресноводные гаплотипы будут сильно отличаться друг от друга. Можно считать, что данные аллели находятся под действием балансирующего отбора в популяции *G. aculeatus* на уровне глобальной метапопуляции [62], так как разные варианты будут полезными в зависимости от конкретной среды обитания. Данная метапопуляция состоит из большой предковой морской и множества потомковых пресноводных популяций. Было показано, что некоторые изменения вследствие адаптации являются специфическими для отдельных популяций [32,63]. Пока плохо изучено, чем отличаются адаптации за счёт новых мутаций и за счёт предсуществующей генетической изменчивости, а также мало изучены различия параллельной адаптации и популяционно-специфических адаптаций.



Рисунок 13. Карты ОД в геноме колюшки, полученные при сравнении морской популяции с несколькими пресноводными популяциями разных возрастов. По вертикальной оси показаны значения F<sub>ST</sub>-статистики, по горизонтальной оси – координаты по длине генома (по [54] с изменениями).

Морские и пресноводные формы *G. aculeatus* имеют некоторое количество фенотипических различий. Наиболее ярким из них является количество костных пластинок на боковой поверхности тела. У морских индивидуумов костные пластины идут от головы до хвоста, а у пресноводных их только немного возле головы и брюшного отдела (рисунок 14, [64,65]). Было показано, что фенотипическое различие в количестве пластин обусловлено 24

различиями в регуляции гена *EDA*, который расположен на хромосоме IV [59,66]. Секвенирование генома *G. aculeatus* способствовало развитию исследований генетики адаптации различных форм колюшки. В результате данных исследований был найден набор геномных участков с повышенной плотностью нуклеотидных замен между морскими и пресноводными индивидуумами при помощи анализа RAD-секвенирования [63]. Спустя два года после этого была опубликована работа, в которой анализировались особи *G. aculeatus* из бассейнов Атлантического и Тихого океанов при помощи полногеномного секвенирования. В данной работе было найдено более 200 ОД, многие из которых являлись неуникальными для каждой отдельной особи [32].



Рисунок 14. Расположение костных пластин вдоль боковой поверхности тела у морских и пресноводных особей трёхиглой колюшки (по [66] с изменениями).

Популяции *G. aculeatus*, адаптированные к пресной воде, населяют водоёмы, образовавшиеся по прошествии последнего ледникового периода, что свидетельствует о быстрой скорости приобретения адаптации [57]. Несмотря на большие различия сред обитания, морские и пресноводные особи обычно могут скрещиваться, образуя плодовитое потомство [56,67,68]. Тем не менее, для некоторых популяций имеет место почти полная репродуктивная изоляция между резидентными (пресноводными) и проходными (морскими) особями, которые могут иметь перекрывающиеся ареалы во время размножения в пресноводных водоёмах [69,70]. В целом нет сильной зависимости между возрастом популяции и степенью репродуктивной изоляции между морскими и пресноводными формами этой популяции.

Исследования особей *G. aculeatus* на Белом море были впервые проведены Валерием Зюгановым в семидесятые годы XX века [71,72]. Верхний предел для возраста морской популяции Белого моря составляет 15.000–18.000 лет, так как ранее данная область была покрыта льдами в течение последнего ледникового периода [73], однако возрасты многих озёрных популяций гораздо меньше. После завершения оледенения происходило эвстатическое повышение уровня воды в Белом море, вследствие чего образовалось большое количество молодых озёр, часть которых окончательно отсоединилась от моря при последующем поднятии поверхности земной коры [74]. Скорость поднятия поверхности земной коры, которая, как показывают наблюдения, всё ещё происходит, примерно постоянна и оценивается ~3,8 мм/год [75]. Данный результат даёт возможность приблизительно рассчитать возрасты пресноводных популяций, исходя из высоты водоёмов над уровнем моря, в которых они обитают. Кроме того, в 1978 г. Валерий Зюганов основал несколько независимых искусственных пресноводных популяций трёхиглой колюшки в заброшенных карьерах, которые были заполнены грунтовыми водами [71]. В каждый из карьеров были помещены равные количества морских и пресноводных особей. В 2011 году из данных популяций были собраны образцы для того, чтобы исследовать эволюционные траектории с известными точками отправления. Таким образом, доступность большого количества молодых озёр с известными возрастами на Белом море предоставляет возможность для изучения динамики адаптации к пресноводной среде обитания. Использование нескольких популяций разных возрастов даёт возможность изучить процесс адаптации в участках под отбором на различных временных промежутках, от десятков до сотен лет, а также исследовать, насколько одинаково происходит действие отбора в независимо адаптирующихся популяциях.

Изучение изменений генома под действием отбора является важным для понимания того, как происходит адаптация на молекулярном уровне. Образование ОД в геноме является следствием действия отбора, а также начальным этапом процесса видообразования, поэтому важно понимать природу и свойства этих структур. Наличие ОД было выявлено во множестве адаптирующихся к разным условиям популяций [34,67,76–78], что свидетельствует в пользу универсальности механизма процесса видообразования у разных видов. Несмотря на распространённость ОД, исследования данных структур ограничены, и до сих пор мало изучены факторы, влияющие на их образование и рост.

## Глава 1. Эволюция локальной скорости мутирования (ЛСМ) и определяющих её факторов

#### 1.1. Материалы и методы

#### 1.1.1. Выравнивания

Было исследовано множественное выравнивание последовательностей ДНК восьми видов приматов: шимпанзе (*P. troglodytes*), гориллы (*G. gorilla gorilla*), орангутанга (*P. pygmaeus abelii*), гиббона (*N. leucogenys*), макаки (*M. mulatta*), зелёной мартышки (*C. sabaeus*), боливийского саймири (*S. boliviensis*) и игрунки (*C. jacchus*) с геномом человека (*H. sapiens*) сборки версии hg19. Данное выравнивание было получено из геномного выравнивания ста позвоночных животных из

UCSC Genome Browser (https://genome.ucsc.edu/; [8]) при помощи программы multiz-tba [79]. Выравнивание было разделено на неперекрывающиеся окна размером 1 Мб или 100 Кб. Экзоны, нетранслируемые области, повторы, непрочитанные позиции (N), ЦпГ-динуклеотиды и гэпы были маскированы. После этого окна, содержащие <20% немаскированных позиций, были удалены из анализа. В результате данной процедуры осталось 2.261 геномных окна длиной 1 Мб и 23.551 геномных окна длиной 100 Кб. Дополнительно также было получено множественное выравнивание тех же девяти геномов приматов с добавлением геномов мыши (*M. musculus*), крысы (*R. norvegicus*) и китайского хомячка (*C. griseus*). В последнем выравнивании наблюдалось значительно больше гэпов, и поэтому были удалены все окна, содержащие <10% немаскированных позиций. После данной процедуры осталось 1.454 геномных окна длиной 1 Мб и 16.449 геномных окна длиной 100 Кб.

Для проверки того, что на полученные результаты не оказывают сильного влияния различия в количестве маскированных нуклеотидов между окнами, отдельные анализы были повторены на окнах, в каждой позиции внутри которых незамаскированные нуклеотиды были у всех видов. Такой анализ был проведён на выравнивании приматов с добавлением мыши, и окна, содержащие <10% позиций после маскирования, были удалены из анализа, после чего осталось 1.091 окон длиной 1 Мб.

#### 1.1.2. Картирование геномных характеристик

Разметки геномных свойств, такие как время репликации, участки гиперчувствительные к ДНКазе, метилирование, СТСF, PolII и модификации гистонов, полученные в результате проекта ENCODE [4], были взяты из геномного браузера UCSC (<u>http://genome.ucsc.edu/ENCODE/</u>). Для большинства анализов были использованы разметки для ткани эмбриональных стволовых клеток, но также в качестве контроля некоторые анализы были повторно сделаны с использованием разметок для других тканей: GM12878, HUVEC, NHEK, Hela-S3 и K562. Карта скорости рекомбинации была взята из проекта HapMap (<u>ftp://ftp.ncbi.nlm.nih.gov/hapmap/</u>). Карта участков, подверженных ГЦ-смещённой генной конверсии, была взята из работы [80] для параметра B=3. Карта распределения нуклеосом была получена из работы [22]. Данные об эволюционной консервативности были взяты из работы [81]. Карта топологических доменов для ткани эмбриональных стволовых клеток была получена из работы [82]. Карта простых повторов была взята из браузера UCSC. Для каждого окна было рассчитано среднее значение каждой карты разметки, исключая маскированные позиции и гэпы. Частота минорного аллеля и данные полиморфизма человеческой популяции были получены для всех ОНП человека за исключением замен W↔S (заменами между парами нуклеотидов Г/Ц и A/ T) при использовании данных проекта 1000 геномов человека [83]. ЛСМ, определённая по полиморфизму (пЛСМ), была посчитана с использованием только 50% наиболее редких ОНП. Список мутаций *de novo* был взят из работы [84].

Карты рекомбинации для шимпанзе и мыши были получены из работ [85,86]. Разметки следующих геномных свойств для ткани эмбриональных стволовых клеток мыши: участки гиперчувствительные к ДНКазе, время репликации и модификации гистонов H3K9me3, H3K27ac, H3K27me3, проекта ENCODE [14] были взяты из геномного браузера UCSC (<u>http://genome.ucsc.edu/ENCODE/</u>). Координаты по геному мыши были конвертированы в координаты по геному человека при помощи программы liftOver [8].

Доля экзонных нуклеотидов была рассчитана для всех нуклеотидов, включая маскированные, в каждом окне. При сравнении изменений геномных разметок между видами значения разметок в окнах были стандартизированы с параметрами среднее = 0 и стандартное отклонение = 1.

#### 1.1.3. Вычисление ЛСМ

Для вычисления ЛСМ были использованы данные по дивергенции (дЛСМ) и по полиморфизму (пЛСМ). Расчёт ЛСМ производился с подсчётом количества однонуклеотидных замен, произошедших в данном окне на данной филогенетической линии при помощи метода «максимальной экономии», как плотность замен среди всех незамаскированных нуклеотидов окна. Для исключения эффектов, связанных с ГЦ-смещённой генной конверсией, из анализа были исключены все замены W↔S. Также для контроля были произведены расчёты на окнах без исключения таких позиций. В качестве альтернативного подхода для вычисления дЛСМ использовалась программа baseml пакета PAML [87], которая использует в своих расчётах метод максимального правдоподобия. Данная программа использовалась с моделью REV при анализе всех типов замен. Результаты альтернативного подсчёта были похожи на полученные в основном анализе (см. ниже).

Аккуратность оценки ЛСМ тестировалась при помощи процедуры бутстрэпа нуклеотидных позиций внутри каждого окна. Рассчитанные пЛСМ и дЛСМ были хорошо скоррелированы со значениями, полученными в результате бутстрэпа (дЛСМ: R=0,97, P<2,2×10<sup>-16</sup>; пЛСМ: R=0,98, P < 2,2×10<sup>-16</sup>; мЛСМ: R=0,75, P < 2,2×10<sup>-16</sup>).

Скорость мутирования различается между разными нуклеотидами [88], и поэтому также проводилась альтернативная процедура по расчёту ЛСМ с учетом оценки вкладов различных

нуклеотидов. Для каждого вида b и геномного окна h было рассчитано ожидаемое количество мутаций *М*<sub>*h,b*</sub> с учётом нуклеотидного состава по следующей формуле:

$$M_{h,b} = \mu_b(\mathbf{G} \leftrightarrow \mathbf{C})n_{h,b}(\mathbf{G} + \mathbf{C}) + \mu_b(\mathbf{A} \leftrightarrow \mathbf{T})n_{h,b}(\mathbf{A} + \mathbf{T})$$

В данной формуле  $\mu_b$  – скорость мутирования для соответствующей замены в виде *b*, а  $n_{h,b}$ - количество сайтов в окне в виде b, где может произойти мутация такого типа. Полученные значения для ожидаемой ЛСМ использовались для нормализации наблюдаемой ЛСМ.

#### 1.1.4. Объясняемая дисперсия ЛСМ

Расчёт дисперсии ЛСМ, объясняемой ЛСМ в близком виде и геномными свойствами, производился при помощи функции линейной регрессии lm в R. Для расчёта независимого вклада в объясняемую дисперсию каждого фактора был использован тест ANOVA III при помощи функции drop1 в R. 95% доверительные интервалы для значений R<sup>2</sup> были получены при помощи процедуры бутстрэпа геномных окон.

#### 1.1.5. Корреляция изменения ЛСМ с изменениями геномных свойств

Для каждого геномного свойства или ЛСМ в каждом окне были посчитаны разницы между соответствующими нормализованными значениями для мыши и человека (а также для шимпанзе в случае рекомбинации). Затем был измерен скорректированный R<sup>2</sup> для корреляций полученных значений для ЛСМ с полученными значениями для каждой геномной характеристики. Также был проведён тест ANOVA III для вычисления независимых вкладов изменений каждой геномной характеристики в изменение ЛСМ. 95% доверительные интервалы для значений R<sup>2</sup> были получены при помощи процедуры бутстрэпа геномных окон.

#### 1.1.6. Поиск участков генома с ускоренной или замедленной ЛСМ

Был произведён поиск геномных окон человека (шимпанзе), для которых ЛСМ была значительно повышена или снижена по сравнению с ЛСМ шимпанзе (человека). Сначала рассчитывалась ожидаемая ЛСМ,  $u_{b,h}(exp)$  в виде *b* для каждого окна *h*, с нормализацией на среднюю ЛСМ для данного окна по всем видам,  $u_h = \frac{\sum_i u_{i,h}}{i}$ , а также с нормализацией на длину филогенетической ветви данного вида,  $v_b = \frac{\sum_j u_{b,j}}{i}$ , и среднюю ЛСМ по всем окнам во всех видах,  $\sum_i u_{i,\underline{h}} \sum_j u_{b,\underline{j}}$ .

$$u_{b,h}(exp) = \frac{\sum_{i} u_{i,h} \sum_{j} u_{b,j}}{\sum_{i} \sum_{j} u_{i,j}}$$

Затем проводилось сравнение полученных ожидаемых значений ЛСМ,  $u_{b,h}(exp)$  человека или шимпанзе, с соответствующими наблюдаемыми значениями ЛСМ,  $u_{b,h}(obs)$ . Для выбора окон, в которых произошли наибольшие изменения ЛСМ в интересующем виде, отбирались те, для которых размер изменения в данном виде  $sp_1$  (человек или шимпанзе) был большим по сравнению с соответствующим изменением в наиболее близком виде  $sp_2$  (соответственно, шимпанзе или человек):

$$\left| lg \frac{u_{sp_1,h}(obs)}{u_{sp_1,h}(exp)} \right| > \left| lg \frac{u_{sp_2,h}(obs)}{u_{sp_2,h}(exp)} \right|$$

Окна ранжировались при помощи статистики, которая показывает, во сколько раз отношение наблюдаемой к ожидаемой ЛСМ в виде *sp*<sub>1</sub> больше соответствующего отношения для вида *sp*<sub>2</sub>:

$$\Delta_h = \frac{u_{sp_1,h}(obs)}{u_{sp_1,h}(exp)} / \frac{u_{sp_2,h}(obs)}{u_{sp_2,h}(exp)},$$

В качестве окон с наиболее ускорившейся или замедлившейся ЛСМ были отобраны 100 окон с самыми высокими или с самыми низкими значениями Δ<sub>h</sub>.

## 1.2. Результаты и обсуждение

#### 1.2.1. ЛСМ сильно скоррелированы между человеком и другими приматами

Для исследования использовалось множественное выравнивание геномов 8 приматов (шимпанзе, горилла, орангутанг, гиббон, макака, зелёная мартышка, боливийский саймири и игрунка) с геномом человека [8], разделённое на 2.261 неперекрывающихся геномных окна длиной 1 Мб. Также для этих окон были получены данные по полиморфизму в человеческой популяции и по *de novo* мутациям из троек семей. Для снижения эффекта действия отбора на оценки ЛСМ из анализа были исключены участки экзонов. Также для оценки силы эффекта отбора был проведён анализ влияния частоты минорного аллеля человеческой популяции на ЛСМ (см. ниже).

Для каждого вида было рассчитано количество замен в его линии относительно другого наиболее близкородственного вида с момента ответвления от последнего общего предка (рисунок 15, Д и Е). Далее полученные количества были поделены на количество незамаскированных нуклеотидов 1 Мб окна (определённая по дивергенции ЛСМ, дЛСМ).

Дополнительно для человека рассчитывалась ЛСМ при использовании ОНП на редких частотах ([83], определённая по полиморфизму ЛСМ, пЛСМ) и ЛСМ, определённая по количеству наблюдаемых в семьях *de novo* мутаций ([84], мЛСМ). мЛСМ является «золотым стандартом» для измерения ЛСМ, но данных для её расчёта мало (рисунок 16), и по этой причине данный метод расчёта ЛСМ использовался только для валидации полученных результатов. мЛСМ немного сильнее скоррелирована с пЛСМ, чем с дЛСМ, хотя обе корреляции значимы (P<0,01; рисунок 17). Из анализа были исключены замены, часть из которых могла возникнуть в результате действия смещённой генной конверсии (см. методы и рисунки 17-18).



Рисунок 15. Объяснение дисперсии ЛСМ в видах приматов при помощи карт геномных свойств и ЛСМ в линии человека. (A– $\Gamma$ ) Для каждого вида приматов доля объясняемой дисперсии в ЛСМ (R<sup>2</sup>, вертикальная ось) изображена в зависимости от филогенетического расстояния этого вида от человека (горизонтальная ось). Значения R<sup>2</sup> для гиббона представлены на графике, но не были включены в регрессию, так как у него наблюдается повышенное по сравнению с другими видами изучаемых приматов количество геномных перестроек [89]. (A–Б) Дисперсия в дЛСМ, объясняемая при помощи дЛСМ (A) или пЛСМ (Б) человека. (B– $\Gamma$ ) Дисперсия в дЛСМ, объясняемая при помощи только геномных свойств человека (красный цвет) или в сочетании с

дЛСМ (В) или пЛСМ (Г; чёрный цвет) человека. Следующие геномные характеристики были включены в модель: ГЦ-состав, скорость рекомбинации, количество экзонных нуклеотидов, время репликации, количество сайтов, чувствительных к ДНКазе, частота минорного аллеля человеческой популяции, плотность меток модификаций гистонов H3K27ac, H3K27me3 и H3K9me3. Окрашенная серым область обозначает часть дисперсии приматов, исключая человека, которая может быть дополнительно объяснена при добавлении в модель ЛСМ человека уже после учёта дисперсии, объясняемой геномными свойствами. Планки погрешностей соответствуют 95% доверительным интервалам, полученным за счёт процедуры бутстрэпа. (Д–Е) Филогенетическое дерево для видов, участвующих в анализе. Красным цветом обозначены ветви, для которых производился расчёт ЛСМ.



Рисунок 16. Корреляции наблюдаемой и полученной в результате процедуры бутстрэпа ЛСМ в 1 Мб геномных окнах. (А) дЛСМ; (Б) пЛСМ; (В) мЛСМ.



Рисунок 17. Дисперсия пЛСМ человека и дЛСМ приматов, объясняемая при помощи мЛСМ человека. (А–Б) W↔S замены не исключены из анализа; (В–Г) W↔S замены исключены из анализа. (А и В) повторы исключены из анализа (такой же набор данных, как и в основных анализах); (Б и Г) повторы не исключены из анализа. При включении участков с повторами было на 63% больше *de novo* мутаций, а также наблюдались более высокие коэффициенты корреляции. (Д–Е) Филогенетическое дерево рассматриваемых в анализе видов. Обозначения, как на рисунке 15.



Рисунок 18. Распределение коэффициентов корреляций между филогенетическим расстоянием от человека до рассматриваемого вида и количеством дисперсии дЛСМ, объясняемой мЛСМ, полученным при процедуре бутстрэпа. (А–Б) W $\leftrightarrow$ S замены включены; (В– $\Gamma$ ) W $\leftrightarrow$ S замены исключены. (А и В) повторы исключены из анализа (такой же набор данных, как и в основном анализе); (Б и Г) повторы не исключены из анализа. Распределения коэффициентов корреляций Спирмена были получены путём 10.000 процедур бустреппа окон. Для каждой панели Р-значения рассчитаны как наименьшие значения для левой и правой квантилей при x=0 с применением поправки на множественное тестирование Бонферрони.

Оценки ЛСМ, полученные при использовании данных по дивергенции и по полиморфизму, могут зависеть от эффектов, связанных со смещённой генной конверсией. Данный эффект заключается в том, что при процессе генной конверсии происходит предпочтительная фиксация сильных (S:  $\Gamma$  или Ц) по сравнению со слабыми (W: A или T) аллелями [27,90,91]. Мы наблюдали, что исключение W $\leftrightarrow$ S замен немного увеличивает корреляцию между дЛСМ и мЛСМ, а также между пЛСМ и мЛСМ (рисунок 17). В целом, достаточно низкая корреляция между пЛСМ и дЛСМ и дЛСМ (рисунок 19, A и B) значительно увеличивается при исключении из анализа W $\leftrightarrow$ S замен (рисунок 15). Примечательно, что чтобы повысить корреляцию между пЛСМ и дЛСМ, а дЛСМ, достаточно исключение W $\leftrightarrow$ S замены только из данных по дивергенции (рисунок 19, Б и  $\Gamma$ ). Последующее исключение W $\leftrightarrow$ S замен из данных по полиморфизму повышает коэффициент

корреляции лишь ненамного, что говорит о том, что смещённая генная конверсия больше влияет на оценку по дЛСМ, чем на оценку по пЛСМ. Во всех основных анализах W↔S замены были исключены.



Рисунок 19. Дисперсия дЛСМ в линиях приматов, объясняемая при помощи пЛСМ и геномных свойств человека. Обозначения те же, что и на рисунке 15. (А и В) W↔S замены включены при оценке дЛСМ и пЛСМ; (Б и Г) W↔S замены исключены при оценке дЛСМ и включены при оценке пЛСМ.

ЛСМ сильно скоррелированы между человеком и шимпанзе ( $R^2$ =0,82 для дЛМР, P<2,2×10<sup>-16</sup>;  $R^2$ =0,46 для пЛСМ, P<2,2×10<sup>-16</sup>; рисунок 15, А и Б). Для менее близкородственных видов данная корреляция падает с ростом филогенетического расстояния, достигая минимального значения среди приматов у игрунки ( $R^2$ =0,27 для дЛСМ, P<2,2×10<sup>-16</sup>;  $R^2$ =0,04 для пЛСМ, P<2,2×10<sup>-16</sup>; рисунок 15, А и Б), и это значение ещё меньше для мыши ( $R^2$ =0,11 для дЛСМ, P<2,2×10<sup>-16</sup>;  $R^2$ =0,02 для пЛСМ, P<3,13×10<sup>-7</sup>; рисунок 20). Данное снижение не зависит от уменьшения выровненных нуклеотидов в окне с ростом филогенетического расстояния, так как коэффициенты корреляций мало изменяются, если в анализе рассматриваются только колонки с немаскированными позициями (рисунок 21). Количество дисперсии ЛСМ, объясняемой ЛСМ человека, уменьшается на половину на филогенетическом расстоянии ~0,04 замены на сайт или ~16 миллионов лет [92], что примерно соответствует последнему общему предку человека и

орангутанга. мЛСМ человека также сильнее скоррелирована с дЛСМ более близких видов, чем с дЛСМ более далёких по филогении видов (рисунок 17).



Рисунок 20. Дисперсия дЛСМ видов приматов и мыши, объясняемая при помощи геномных свойств и дЛСМ (А и В) или пЛСМ (Б и Г) человека. Обозначения такие же, как и на рисунке 15. Набор видов соответствует набору видов на рисунке 15 с добавлением точки для мыши (крайняя правая точка).


Рисунок 21. Дисперсия ЛСМ приматов и мыши, объясняемая при помощи геномных свойств и дЛСМ (А и В) или пЛСМ (Б и Г) человека для выравнивания с исключёнными колонками, в которых хотя бы один из рассматриваемых видов содержал замаскированную позицию. Обозначения такие же, как и на рисунке 15. Набор видов соответствует набору видов на рисунке 15 с добавлением мыши (крайняя правая точка).

ЛСМ зависит от свойств ДНК [6,29]. Линейная модель, предсказывающая дЛСМ человека при помощи разметок геномных свойств ткани эмбриональных стволовых клеток, объясняет только 33% дисперсии дЛСМ, что немного выше оценки полученной в предыдущей работе [6] при использовании разметок другой ткани (см. ниже). Для каждого вида кроме человека была сделана модель при использовании только геномных свойств, а также в сочетании с дЛСМ или пЛСМ человека. дЛСМ приматов может быть предсказана геномными свойствами почти также хорошо, как и дЛСМ человека, что соотносится с тем, что распределения геномных свойств достаточно консервативны этими видами. Также большая часть исследованных свойств имеет значимую корреляцию между человеком и мышью (таблица 1).

 Геномное свойство	<b>R</b> Пирсона	Р-значение	
 ЛСМ	0,34 (0,29; 0,38)	<2,2×10 <sup>-16</sup>	
Скорость рекомбинации	0,01 (-0,04; 0,06)	0,67	
ДНКаза	0,83 (0,82; 0,85)	<2,2×10 <sup>-16</sup>	
Время репликации	0,71 (0,69; 0,74)	<2,2×10 <sup>-16</sup>	
ГЦ-состав	0,94 (0,93; 0,94)	<2,2×10 <sup>-16</sup>	
Плотность экзон. нукл.	0,95 (0,94; 0,95)	<2,2×10 <sup>-16</sup>	
H3K9me3	0,32 (0,28; 0,37)	<2,2×10 <sup>-16</sup>	
H3K27ac	0,73 (0,70; 0,75)	<2,2×10 <sup>-16</sup>	
H3K27me3	0,66 (0,63; 0,69)	<2,2×10 <sup>-16</sup>	

Таблица 1. Коэффициенты корреляций Пирсона между геномными свойствами человека и мыши по 1 Мб геномным окнам. 95% доверительные интервалы (асимптотические доверительные интервалы, оцененные при помощи Z-трансформации Фишера) показаны в скобках. ДНКаза – плотность участков, чувствительных к ДНКазе.

Добавление ЛСМ в модель позволяет значительно повысить количество объясняемой дисперсии (рисунок 15В–Г): от R<sup>2</sup>=0,42 до 0,72 (для пЛСМ) или до 0,84 (для дЛСМ) для шимпанзе, и от 0,40 до 0,71 (для пЛСМ) или до 0,80 (для дЛСМ) для гориллы. Таким образом, для близких видов семейства *Hominidae* линейная модель, которая включает данные геномных свойств и данные, полученные при использовании человеческого полиморфизма или дивергенции, позволяет объяснить до ~2 раз больше дисперсии ЛСМ по сравнению с моделью, включающую только геномные свойства. Остальная часть дисперсии (рисунок 15 В–Г, серая область) может быть случайной или же быть связана с геномными свойствами, не использованными в данном анализе. Часть дисперсии, которая не объясняется использованными геномными свойствами, вероятно, содержит компоненту, консервативную между близкими видами.

Анализ корреляции дЛСМ между видами также был проведён на геномных окнах 100 Кб. После процедуры фильтрации окон всего осталось 23.551 геномных окон длиной 100 Кб для выравнивания приматов и 16.449 геномных окон длиной 100 Кб для выравнивания приматов с добавлением трёх видов грызунов. Корреляция между дЛСМ человека и шимпанзе в 100 Кб окнах ( $R^2$ =0,36, P<2,2×10<sup>-16</sup>; рисунок 22) была ниже наблюдаемой для 1 Мб окон, что частично может объясняться меньшей зависимостью от локальных эффектов мелкого масштаба более длинных окон. Снижение значений коэффициентов корреляций с ростом филогенетического

расстояния было схожим с наблюдаемым для 1 Мб окон: наименьшее значение объяснённой дисперсии среди приматов было для игрунки ( $R^2=0,11$ ,  $P<2,2\times10^{-16}$ , рисунок 22A), и это значение было ещё меньше для мыши ( $R^2=0,06$ ,  $P<2,2\times10^{-16}$  для 100 Кб окон; рисунок 22Б).



Рисунок 22. Объяснение дисперсии ЛСМ у видов приматов (А) и у видов приматов с добавлением мыши (Б) при помощи ЛСМ человека в 100 Кб геномных окнах. Обозначения те же, что и на рисунке 15.

Несмотря на то, что мЛСМ является наиболее прямой оценкой ЛСМ, мЛСМ оценки более смещены по сравнению с теми, что получены по пЛСМ и дЛСМ вследствие того, что набор мутаций, которые используются при оценке мЛСМ, недостаточно большой. Так, для человека при расчёте ЛСМ наш анализ включает 636.954 мутаций для дЛСМ, 1.093.223 мутаций для пЛСМ, но только 11.429 мутаций для мЛСМ. Вследствие этого ожидаются более низкие коэффициенты корреляций между мЛСМ и дЛСМ, а также между мЛСМ и пЛСМ даже для одного и того же вида по причине малого числа событий мутаций.

Для проверки данного предположения отбирались подмножества из общего набора позиций замен (по которым рассчитывался дЛСМ) и полиморфных событий (по которым рассчитывался пЛСМ), и проверялось, насколько хорошо подмножества событий объясняют дисперсию полных наборов событий. При использовании меньших подмножеств событий получались меньшие значения  $R^2$  (рисунок 23). Так, при корреляции данных по 11.429 событий мутаций против полного набора событий дЛСМ или пЛСМ получался  $R^2$ ~0,24 в обоих случаях. Поэтому низкая корреляция между мЛСМ и дЛСМ, а также между мЛСМ и пЛСМ, даже в одном и том же виде может быть отчасти объяснена недостаточным количеством событий *de novo* мутаций для оценки ЛСМ.



Рисунок 23. Корреляция оценки ЛСМ человека с той же оценкой, но полученной по меньшему набору событий. (А) дЛСМ; (Б) пЛСМ. Горизонтальная ось показывает долю, которую поднабор мутаций составляет от всего набора; вертикальная ось соответствует доле объяснённой дисперсии (R<sup>2</sup>). Штрихованная линия пересекает горизонтальную ось при значении, которое соответствует 11.429 событиям мутаций. Красными кругами обозначены R<sup>2</sup> для корреляций мЛСМ и дЛСМ (А) или пЛСМ (Б).

В качестве альтернативного подхода использовался расчёт дЛСМ при помощи метода максимального правдоподобия программой baseml с моделью REV (рисунок 24). Также для контроля при расчёте скоростей мутирования использовались разные скорости мутирования для разных замен (см. методы). Результаты получились схожими с теми, что были получены без данной нормализации (рисунок 25).



Рисунок 24. Дисперсия ЛСМ, объясняемая при помощи геномных свойств и дЛСМ человека. дЛСМ были рассчитаны при помощи метода максимального правдоподобия и без исключения W↔S замен. Обозначения такие же, как на рисунке 15.



Рисунок 25. Дисперсия ЛСМ, объясняемая при помощи геномных свойств и дЛСМ (А и В) или пЛСМ (Б и Г) человека. дЛСМ были рассчитаны при учёте различий в скоростях мутирования между заменами и видами. Обозначения такие же, как на рисунке 15.

### 1.2.2. Объяснение дисперсии ЛСМ при помощи геномных свойств

ЛСМ в геномных участках изменяется с течением времени (рисунок 15). Количество дисперсии ЛСМ, объясняемой при помощи человеческих геномных свойств, является схожим как для близкородственных видов, так и для более далёких на филогении приматов. Тем не менее есть вероятность, что отдельные геномные свойства, а также их способность предсказывать ЛСМ, могут меняться на других временных промежутках и эти эффекты будут заметны при сравнении более давно разошедшихся по филогении видов. Мы исследовали, насколько изменения в ЛСМ зависят от изменений геномных свойств. Данные для геномных свойств приматов имеются лишь в ограниченном количестве [93,94], и поэтому для исследования использовался особый подход.

Вначале оценивалось количество дисперсии ЛСМ, объясняемой отдельными геномными свойствами. Известно, что геномные свойства скоррелированы друг с другом ([6], рисунок 26 и таблица 1). Поэтому для выяснения вкладов геномных свойств в объяснение дисперсии ЛСМ, независимых от других геномных свойств, был проведён анализ ANOVA III типа (рисунок 27). Полученные оценки для вкладов геномных свойств в предсказание дЛСМ человека согласуются

с результатами предыдущих исследований [6,7,29]. Вклады геномных свойств в предсказание дЛСМ остальных приматов являются схожими и для большинства свойств не зависят от филогенетического расстояния от человека до рассматриваемого вида (рисунок 27). Единственной из рассматриваемых геномных характеристик, для которой вклад в объяснение дисперсии снижается с ростом филогенетического расстояния, является скорость рекомбинации (P=0.009 для корреляции между R<sup>2</sup> и филогенетическим расстоянием, рисунки 27 и 28).Количество дисперсии, объясняемой данной характеристикой, является более высоким для более близких видов, что соотносится с наблюдением о её большом влиянии на ЛСМ [95,96]. Тем не менее, доля объясняемой рекомбинацией дисперсии очень быстро снижается с ростом филогенетического расстояния (от 6,05% для человека до 0,01% для игрунки, рисунки 27 и 28). Данное снижение, вероятно, связано непосредственно с рекомбинацией, а не со связанным с ней процессом генной конверсии, так как в нашем анализе были исключены все замены, которые могут получаться в результате смещённой генной конверсии. Полученные результаты согласуются с наблюдениями влияния рекомбинации на ЛСМ, независимого от смещённой генной конверсии [16,97]. Наблюдаемое снижение объясняемой рекомбинацией доли дисперсии ЛСМ даёт основания предположить, что изменения скорости рекомбинации происходят быстро, что соответствует её известной способности к быстрой эволюции [18,85,90]. Такие изменения способны влиять на изменения ЛСМ.



Рисунок 26. Модули коэффициентов попарных корреляций между всеми парами геномных характеристик и ЛСМ человека. Более тёплые цвета соответствуют более высоким абсолютным значениям коэффициентов корреляций Пирсона.



Рисунок 27. Дисперсия ЛСМ, объясняемая 25 геномными свойствами. Для каждого геномного свойства количество объясняемой дисперсии ЛСМ, определённой при помощи анализа ANOVA III, показано для каждого вида приматов (виды расположены для каждого геномного свойства слева направо – с увеличением филогенетического расстояния от человека, как на дереве рисунка 15). Геномные характеристики расположены в порядке уменьшения объясняемой ими дисперсии в линии человека. Звёздочками обозначена значимость отрицательной корреляции между R<sup>2</sup> и филогенетическим расстоянием.



Рисунок 28. Распределения коэффициентов корреляций между филогенетическим расстоянием от человека и количеством дисперсии дЛСМ, объясняемой каждым геномным свойством (анализ ANOVA тип III). Распределения коэффициентов корреляций Спирмена были получены при помощи процедуры бутстрэпа 1 Мб окон 10.000 раз. Для каждого геномного свойства Р-значения рассчитаны как наименьшие значения для левой и правой квантилей при х=0 с применением поправки на множественное тестирование Бонферрони.

В основных анализах использовались разметки линии эмбриональных стволовых клеток. Для контроля анализ вкладов геномных свойств в объяснение дисперсии ЛСМ был произведён также при использовании карт геномных характеристик для нескольких других тканей. В предыдущей работе [6] при использовании разметок линии GM12878 было объяснено 28% дисперсии ЛСМ человека. В анализе данной работы было объяснено 36% дисперсии ЛСМ человека при использовании меньшего набора геномных свойств данной ткани (таблица 2). Наблюдаемая разница, вероятно, является следствием того, что в анализах использовались различные сборки генома человека (hg18 в [6] и hg19 в нашем анализе), а также различные свойства используемых выравниваний (данные по дивергенции были получены из выравнивания с использованием ЕРО алгоритма в [6], а в нашем исследовании было использовано выравнивание из геномного браузера UCSC) и процедуры маскирования и расчёта ЛСМ.

Дополнительно был проведён анализ ANOVA тип III для выявления индивидуальных вкладов геномных характеристик в объяснение дисперсии ЛСМ для разных тканей (рисунок 29). Результаты были очень схожи с теми, что были получены в основном анализе (таблица 2, рисунок 27). В частности, для всех тканей наблюдалось снижение вклада скорости рекомбинации в объяснение дисперсии ЛСМ с ростом филогенетического расстояния.

	Человек	Шимпанзе	Горилла	Оранг.	Гиббон	Макака	Зелёная	Болив.	Игрунка
							мартышка	саймири	
H1Esc	0,33	0,37	0,36	0,38	0,26	0,38	0,31	0,39	0,46
K562	0,32	0,35	0,33	0,35	0,23	0,34	0,28	0,35	0,42
NHEK	0,37	0,40	0,39	0,43	0,30	0,41	0,32	0,40	0,47
GM12878	0,36	0,39	0,38	0,42	0,28	0,39	0,30	0,40	0,47
HelaS3	0,33	0,35	0,34	0,34	0,24	0,31	0,26	0,35	0,40
Huvec	0,37	0,39	0,38	0,43	0,29	0,39	0,31	0,41	0,47

Таблица 2. Доля дисперсии дЛСМ, объясняемая геномными свойствами, полученными по разным тканям. Девять геномных свойств соответствует набору, который был использован при построении рисунка 15. Только пять из девяти геномных свойств варьировали между рассматриваемыми тканями: время репликации, карта участков, гиперчувствительных к ДНКазе, и карты распределений трёх разметок гистонов.



филогенетическое расстояние



филогенетическое расстояние

Рисунок 29. Дисперсия ЛСМ, объясняемая геномными свойствами, полученными по разным тканям. (А) GM12878; (Б) HUVEC; (В) NHEK; (Г) Hela-S3; (Д) K562. Обозначения такие же, как на рисунке 27.

## 1.2.3. Изменения скорости рекомбинации связаны с изменениями ЛСМ

Далее была произведена оценка влияния изменений скоростей рекомбинации на эволюцию ЛСМ между человеком и шимпанзе или человеком и мышью. Для шимпанзе и мыши карты рекомбинации были взяты из работ [85,86]. Было найдено, что рекомбинация сильно различается между двумя видами: скорость рекомбинации человека только слабо скоррелирована со скоростью рекомбинации шимпанзе ( $R^2$ =0,24, P<2,2×10<sup>-16</sup>) и совсем не скоррелирована с соответствующей у мыши (таблица 1).

Было проведено сравнение разницы в дЛСМ с разницей в скорости рекомбинации для каждого 1 Мб геномного окна. В обоих сравнениях (человек-шимпанзе, человек-мышь) данные факторы имели слабую положительную корреляцию ( $R^2=0,01$ ,  $P<7,0\times10^{-6}$  для шимпанзе, рисунок 30А; и  $R^2=0,1$ ,  $P<2,2\times10^{-16}$  для мыши, рисунок 30Б). Данный результат показывает, что повышение скорости рекомбинации в геномном участке связано с повышением его ЛСМ и наоборот.

Для сравнения человек-мышь были проанализированы также и другие геномные характеристики на предмет ассоциации их изменений в распределении между видами с изменениями в ЛСМ. В целом ~16,4% всей дисперсии изменения ЛСМ между видами может быть объяснено изменениями геномных свойств (рисунок 30). Когда рассматривались отдельные вклады геномных характеристик в объясняемую дисперсию, было выявлено, что изменение скорости рекомбинации больше других свойств объясняет дисперсию изменения ЛСМ (~10%, рисунок 30, А-Б и К). Для выделения независимых вкладов изменений в геномных свойствах между человеком и мышью в объяснение дисперсии изменений ЛСМ между этими двумя видами был использован анализ ANOVA III (рисунок 30Л). Изменения только нескольких геномных свойств вносили существенный вклад в изменения ЛСМ. Наибольшее количество дисперсии было объяснено изменением в скорости рекомбинации; вклад этого свойства был гораздо больше, чем следующего по значению фактора (ГЦ-состава). Полученный результат показывает, что, несмотря на быструю эволюцию распределения горячих точек рекомбинации и то, что карта рекомбинации человека нисколько не предсказывает карту рекомбинации мыши (таблица 1), изменения рекомбинации, которые происходят между этими двумя видами, объясняют изменения в ЛСМ сильно лучше, чем изменения других геномных свойств. Также в качестве контроля аналогичный анализ был проведён на данных при учёте различий в скоростях мутирования между разными нуклеотидами (рисунок 31).



Рисунок 30. (А-И) Графики корреляций между изменениями ЛСМ и изменениями некоторых геномных свойств между видами человек и мышь (или шимпанзе в случае рекомбинации). Каждая точка соответствует одному 1 Мб геномному окну. (К-Л) Изменения в ЛСМ, объясняемые изменениями в некоторых геномных свойствах между человеком и мышью. Вертикальная ось соответствует количеству дисперсии в изменении ЛСМ, объясняемой изменениями геномных свойств между человеком и мышью. Каждый столбец соответствует количеству дисперсии, объясняемой конкретной геномной характеристикой, посчитанной через  $R^2$  (К) или по анализу ANOVA III (Л). Планки погрешностей соответствуют 95% доверительным интервалам, полученных количеств объяснённой дисперсии (\*: P<0,05, \*\*: P<0,01, \*\*\*: P<0,001).



Рисунок 31. ЛСМ, объясняемая отдельными геномными свойствами. дЛСМ были рассчитаны при учёте различий в скоростях мутирования между разными нуклеотидами (см. методы). Обозначения такие же, как на рисунках 27 и 30.

Для того чтобы лучше разобраться в связи изменений скоростей рекомбинации с изменениями в ЛСМ, был проведён анализ геномных окон, для которых значения ЛСМ были значительно повышены или снижены в человеке или шимпанзе с момента их дивергенции от общего предка человека и шимпанзе (таблица 3). В окнах с ускорившейся ЛСМ в линии человека скорость рекомбинации человека была значительно выше соответствующего среднего значения по геному (односторонний тест Манна-Уитни  $P=2,9\times10^{-9}$ , рисунок 32A), и это говорит о том, что такие окна часто содержат в себе горячие точки рекомбинации. Напротив, в окнах, для которых значение ЛСМ повышено у шимпанзе, скорость рекомбинации человека не так сильно повышена по сравнению со средним значением по геному ( $P=8,0\times10^{-3}$ , рисунок 32Б). Полученные результаты дают основания предположить, что окна с повышенной ЛСМ в человеке часто связаны с присутствием горячих точек рекомбинации, которые являются короткоживущими и поэтому повышают ЛСМ в человеке гораздо сильнее, чем в шимпанзе. Аналогично, окна с повышенной ЛСМ в шимпанзе также имеют повышенную скорость рекомбинации ( $P=3,7\times10^{-7}$ , рисунок 32Г), немного превышающую соответствующее значение в окнах с ускорившейся ЛСМ в человеке  $Q=7,4\times10^{-6}$ , рисунок 32В).

	зам. в Ч.	зам. в Ш.	среднее по	уск. в Ч.	уск. в Ш.
			геному		
Человек	0,00058 (79%)	0,00070 (95%)	0,00074	0,00104 (141%)	0,00076 (103%)
Шимпанзе	0,00080 (97%)	0,00066 (79%)	0,00083	0,00098 (118%)	0,0010 (122%)

Таблица 3. Средние ЛСМ человека и шимпанзе в участках с замедленной ЛСМ в человеке (зам. в Ч.) и шимпанзе (зам. в Ш.), а также с ускоренной ЛСМ в человеке (уск. в Ч.) и шимпанзе (уск. в Ш.). Значения в скобках соответствуют процентным соотношениям относительно средних значений по геному.



Рисунок 32. Распределения значений скоростей рекомбинации человека (А–Б) в окнах с ускоренной ЛСМ в человеке (уск. в Ч.; А) и шимпанзе (уск. в Ш.; Б), и распределение значений скоростей рекомбинации шимпанзе (В–Г) в окнах с ускоренной ЛСМ в человеке (В) и шимпанзе (Г). Схематические филогении показывают линию вида, в котором ЛСМ была повышена, красным цветом (Ч, человек или Ш, шимпанзе); виды, в которых измерялась рекомбинация, обозначены кругами. Штрихованная линия отмечает медианное значение скорости рекомбинации в участках с ускоренной ЛСМ и по всему геному.

Также был проведён анализ для окон с замедлившейся ЛСМ в человеке или шимпанзе. Скорость рекомбинации человека была немного снижена по сравнению со средним значением в окнах с замедлившейся ЛСМ как человека, так и шимпанзе (односторонний тест Манна-Уитни  $P=7,7\times10^{-4}$  и  $P=3,9\times10^{-4}$  соответственно; рисунок 33). Данное наблюдение даёт основания предполагать, что снижение ЛСМ также может частично зависеть от факторов, связанных со скоростью рекомбинации, но данный эффект проявляется более слабо, чем для окон с повышенной ЛСМ. Все эти наблюдения, как показывают анализы, связаны непосредственно с рекомбинацией, а не со связанным фактором смещённой генной конверсии (см. ниже).



Рисунок 33. Распределения значений скоростей рекомбинации человека (А–Б) в окнах с замедленной ЛСМ в человеке (зам. в Ч.; А) и в шимпанзе (зам. в Ш.; Б), и распределение значений скоростей рекомбинации шимпанзе (В–Г) в окнах с замедленной ЛСМ в человеке (В) и в шимпанзе (Г). Схематические филогении показывают линию вида, в котором ЛСМ была снижена, синим цветом (Ч, человек или Ш, шимпанзе); виды, в которых измерялась рекомбинация, обозначены кругами. Штрихованная линия отмечает медианное значение скорости рекомбинации в участках со сниженной ЛСМ и по всему геному.

Известно, что ГЦ-смещённая конверсия ассоциирована с рекомбинацией. Поэтому для анализа её индивидуального вклада в объяснение дисперсии ЛСМ дополнительно были использованы данные по скорости генной конверсии из работы [80]. Значения смещённой генной конверсии человека были значимо выше в участках с ускоренной ЛСМ в линии человека и шимпанзе по сравнению со средним значением по геному (односторонний тест Манна-Уитни, P=7,34×10<sup>-9</sup> и P=0,04, соответственно для человека и шимпанзе; рисунок 34, А и В). Также значения генной конверсии были существенно снижены в участках с замедленной ЛСМ в линии человека и шимпанзе (P=0,007 и P=0,006, соответственно; рисунок 34, Б и Г). Таким образом, данные по рекомбинации и генной конверсии подтверждают роль специфических для вида горячих точек рекомбинации в повышении ЛСМ.



Рисунок 34. Распределения значений скоростей смещённой генной конверсии человека в окнах с ускоренной ЛСМ в человеке (уск. в Ч.; А), замедленной ЛСМ в человеке (зам. в Ч.; Б), ускоренной ЛСМ в шимпанзе (уск. в Ш.; В) и замедленной ЛСМ в шимпанзе (зам. в Ш.; Г). На схемах филогений показана линия, в которой ЛСМ была повышена (красный цвет) или понижена (синий цвет), а также окружностью обозначен вид, для которого измерялась скорость смещённой генной конверсии (человек). Штрихованная линия соответствует медиане значений скоростей смещённой генной конверсии в окнах с повышенной или пониженной ЛСМ, а также по всему геному.

Наблюдаемая связь между скоростью рекомбинации и ЛСМ может возникать за счёт непосредственного эффекта смещённой генной конверсии [27], которая благоприятствует закреплению W→S мутаций. Тем не менее, эффект не объясняется полностью повышенными значениями генной конверсии, так как эффект остаётся и после исключения нуклеотидных замен W→S и S→W. Данный результат был получен в обоих случаях, когда использовались значения скоростей рекомбинации (рисунки 32-33 и 35-36) и генной конверсии (рисунки 34 и 37).



Рисунок 35. Распределения значений скоростей рекомбинации человека (А–Б) в окнах с ускоренной ЛСМ в человеке (уск. в Ч.; А) и шимпанзе (уск. в Ш.; Б), и распределение значений скоростей рекомбинации шимпанзе (В–Г) в окнах с ускоренной ЛСМ в человеке (В) и шимпанзе (Г). Анализ проводился на данных без исключения W↔S замен. Обозначения такие же, как и на рисунке 32.



Рисунок 36. Распределения значений скоростей рекомбинации человека (A–Б) в окнах с замедленной ЛСМ в человеке (зам. в Ч.; А) и в шимпанзе (зам. в Ш.; Б), и распределение значений скоростей рекомбинации шимпанзе (B– $\Gamma$ ) в окнах с замедленной ЛСМ в человеке (B) и в шимпанзе ( $\Gamma$ ). Анализ проводился на данных без исключения W $\leftrightarrow$ S замен. Обозначения такие же, как и на рисунке 32.



Рисунок 37. Распределения значений скоростей смещённой генной конверсии человека в окнах с ускоренной ЛСМ в человеке (уск. в Ч.; А), замедленной ЛСМ в человеке (зам. в Ч.; Б), ускоренной ЛСМ в шимпанзе (уск. в Ш.; В) и замедленной ЛСМ в шимпанзе (зам. в Ш.; Г). Анализ проводился на данных без исключения W↔S замен. Обозначения те же, что и на рисунке 34.

## 1.3. Выводы

Известно, что ЛСМ может варьировать очень сильно внутри одного генома, но причины такой изменчивости остаются не до конца понятны. В данной работе было показано, что предсказание ЛСМ значительно улучшается при использовании ЛСМ близкородственных видов. Также было установлено, что схожесть ЛСМ между видами является более короткоживущей, чем сходство большинства геномных разметок, которые показывают неплохую корреляцию даже для таких эволюционно далёких видов, как человек и мышь. Тем не менее, не все геномные разметки сохраняют свое распределение на таких больших временах. Одной из наиболее быстро эволюционирующих разметок является скорость рекомбинации, которая достаточно сильно различается уже между человеком и шимпанзе, и корреляция которой между гомологичными участками пропадает полностью при сравнении карт рекомбинации человека и мыши. Было показано, что скорость рекомбинации является фактором, индивидуальный вклад которого изменяется сильнее других на филогении приматов, и что данный фактор ассоциирован с сегментами генома с ускоренной или замедленной ЛСМ.

Среди факторов, определяющих ЛСМ, использовались также те, что связаны с действием отбора, такие как частота минорного аллеля, плотность экзонных нуклеотидов и консервативность. Данные факторы находятся среди геномных свойств, вносящих наибольший вклад в объяснение ЛСМ человека. Тем не менее, обычно для детектирования неравномерности действия отбора используются геномные окна меньших размеров вследствие того, что отбор чаще действует на отдельные гены или регуляторные элементы. Результатами действия положительного отбора являются участки генома с локально повышенной межпопуляционной дивергенцией, изучение которых позволяет делать выводы об истории их возникновения и о процессах, происходящих в популяции.

# Глава 2. Геномная архитектура адаптации у трёхиглой колюшки

# 2.1. Материалы и методы

## 2.1.1. Сбор образцов и геномное секвенирование

Сбор образцов популяций трёхиглой колюшки производился в июне-августе 2011 г. при помощи сачка. Описание популяций, включая расположение мест сбора, возрасты популяций и количество особей в каждом образце, приведены в таблице 4. ДНК была экстрагирована из каждого индивидуума при использовании Wizard genomic DNA purification kit (Promega). Перед приготовлением библиотек ДНК для каждого объединенного образца, содержащего от 8 до 20 индивидуумов, смешивались в равных соотношениях. Полученные образцы ДНК были обработаны в соответствии с протоколом TruSeq DNA Sample Preparation Guide (Illumina). Библиотеки подвергались флюориметрии Qubit (Invitrogen) и ПЦР реального времени (праймеры I-qPCR-1.1: AATGATACGGCGACCACCGAGAT I-qPCR-2.1: И СААССАСААСАСССАТАССА) и затем разбавлялись до окончательной концентрации 9 рМ. Полученные библиотеки были секвенированы на HiSeq2000 при помощи набора TruSeq SBS Kit v3-HS (Illumina) с длиной чтений 101 нуклеотид. Чтения для каждой из популяций были выложены в архив NCBI Short Read Archive (http://www.ncbi.nlm.nih.gov/Traces/sra; идентификационный номер проекта SRP023197), где находятся в свободном доступе.

Образец	Описание	Координаты	Кол-во	Среднее	Нукл.
		места сбора	особей	покрытие	разнообр.
					(π)
Нильма <sup>а</sup>	Морские	66° 30.45'N,	16	44	0.00210
		33°7.68'E			
Ершовское	Проходные, собранные в	66°32.21'N,	10	17	0.00188
(проходные) <sup>а</sup>	Ершовском	33°3.62'E			
Ершовское	Пресноводные,	66°32.21'N,	12	52	0.00219
(резидентные) <sup>ь</sup>	возраст ~34 г. [32]	33°3.62'E			
Марцы <sup>ь</sup>	Пресноводные,	66°35.95'N,	10	24	0.00181
	возраст ~250 л.	33°15.35'E			
Голубой <sup>с</sup>	Карьер, популяция заселена	66°17.20'N,	20	48	0.00204
	в 1978 г. 20-ю морскими и	33°23.29'E			
	20-ю пресноводными				
	рыбами [33]				
Малыш <sup>с</sup>	Карьер, популяция заселена	66°18.27'N,	20	63	0.00186
	в 1978 г. 1-й морской и 1-й	33°25.27'E			
	пресноводной рыбами [33]				
Лобанешское <sup>d</sup>	Пресноводные, ~600 л.	66°33.64'N,	8	23	0.00167
		33°13.45'E			
Машинное <sup>d</sup>	Пресноводные, ~700 л.	66°17.74'N,	10	19	0.00158
		33°21.82'E			

Таблица 4. Места сбора и характеристики образцов. <sup>а</sup>морская популяция; <sup>b</sup>молодые природные пресноводные популяции; <sup>с</sup>искусственные пресноводные популяции; <sup>d</sup>более старые природные пресноводные популяции.

### 2.1.2. Геномное картирование и анализ последовательностей

Геномные чтения были картированы на референсный геном вида G. aculeatus, полученный из геномного браузера UCSC (http://genome.ucsc.edu/), при помощи программы bwa aln пакета **BWA** (Burrows-Wheeler http://bio-bwa.sourceforge.net/). Alignment Tool: Полученное выравнивание затем было конвертировано в SAM-формат при помощи программы bwa sampe. Далее данные были обработаны при помощи программы picard (http://picard.sourceforge.net/) для удаления чтений-дубликатов. Индивидуальные для каждой популяции ОНП были идентифицированы помощи программы mpileup samtools при пакета (http://samtools.sourceforge.net/). Затем для полученных для каждого образца наборов ОНП проводилась процедура фильтрации по покрытию. В качестве порогов были взяты следующие значения покрытий на позицию: >10 для популяций Нильма, Малыш, Голубой и пресноводных особей из Ершовского, и >5 для образцов из озёр Машинное, Лобанешское, Марцы и проходных особей из Ершовского. Чтобы уменьшить влияние на результаты ошибок секвенирования, были исключены все позиции с качеством <40. В качестве альтернативного подхода идентификации ОНП была программа UnifiedGenotyper GATK использована ИЗ пакета (http://www.broadinstitute.org/gatk/) с теми же порогами по минимальному покрытию. Данная процедура привела к получению большего количества ОНП, но были получены похожие паттерны кластеризации. Позиции генов были получены из аннотации Ensembl версии 72 (http://www.ensembl.org/).

## 2.1.3. Кластеры маркерных однонуклеотидных полиморфизмов в геноме

Частоты пресноводных аллелей рассчитывались для каждого образца, и для некоторых ОД эти оценки валидировались при помощи аллель-специфического ПЦР (см. ниже). Маркерный ОНП определялся как полиморфная позиция, в которой обе морские популяции содержали один и тот же аллель на частоте ≥80%, в то время как обе пресноводные популяции содержали другой аллель с частотой ≥80% («сильный критерий») или же на частоте ≥50%, что соответствовало «слабому критерию». Для сравнения также определялись маркерные ОНП по сильному критерию при помощи только одной морской и одной пресноводной популяций (из Нильмы и озера Машинного).

При поиске ОД объединялись все найденные 10 Кб геномные окна с количеством маркерных ОНП >20, которые находились на расстоянии менее 40 Кб друг от друга. В целом данная процедура помогала избегать увеличения количества ОД из-за недостаточного покрытия отдельных участков генома. Без объединения близкорасположенных ОД значительно увеличивалось количество ОД (71), но качественные особенности динамики частот

пресноводных аллелей между популяциями оставались схожими. Маловероятно, что процедура объединения оказывает сильное влияние на полученные эффекты, поскольку вероятность рекомбинации за время жизни рассматриваемых популяций для участков генома, находящихся на расстоянии объединения, достаточно низка (средняя по геному скорость рекомбинации у трёхиглой колюшки составляет 3,11 См [98]).

#### 2.1.4. Валидация частот пресноводных аллелей в островах дивергенции (ОД)

Для валидации наших оценок частот пресноводных аллелей, полученных при секвенировании, были генотипированы особи, использованные при приготовлении образцов ДНК популяций. Генотипирование проводилось при помощи набора аллель-специфических праймеров для участков, входящих в состав некоторых из ОД. Особи с одним или двумя ПЦР продуктами были идентифицированы соответственно, как гомо- и гетерозиготы. Полученные частоты пресноводных аллелей хорошо соответствовали оценкам при использовании данных геномного секвенирования (см. ниже).

#### 2.1.5. Оценка коэффициентов отбора в ОД

Для каждого ОД была рассчитана средняя частота пресноводного аллеля по всем маркерным ОНП, находящимся в пределах ОД. Средняя длина поколения использовалась в расчётах равной 2 годам [99]. Особи более старших возрастов также способны к размножению, и в таком случае мы будем несколько недооценивать коэффициенты отбора *s*. Возрасты искусственных популяций карьеров Голубой и Малыш известны равными 34 годам (17 поколений). Возраст озера Ершовского по данным исследования популяции составляет 34 года или меньше [71], и он принимался в расчётах равным 34 годам (17 поколений).

Изначальные частоты пресноводных аллелей в двух искусственных популяциях Голубой и Малыш полагались равными 0,5, так как данные популяции основывались одинаковыми количествами морских и пресноводных особей. В 1978 году все особи озера Ершовское имели фенотипические признаки, схожие с наблюдающимися у морских особей [71]. Исходя из этого, в своих расчётах мы предполагали, что изначальные частоты пресноводного аллеля в популяции озера Ершовское соответствовали наблюдаемым в морской популяции, то есть  $p_0~0,1$ . Коэффициент отбора *s* рассчитывался при помощи уравнения изменения частоты аллеля с каждым поколением при предположении, что на динамику оказывает влияние только отбор (уравнение 3.2 из [100], при использовании параметра для неполного доминирования h=0,5).

## 2.2. Результаты и обсуждение

### 2.2.1. Генетические различия между морской и пресноводными популяциями

Был произведён поиск различий между предковой популяцией G. aculeatus из Белого моря и потомковыми пресноводными популяциями озёр бассейна Белого моря. Для поиска таких маркеров сравнивались геномные последовательности двух образцов морской популяции с двумя образцами пресноводных популяций (рисунок 38, таблица 4). Морские особи собирались в заливе Нильма, а также в озере Ершовском, куда они приходят на нерест. Принадлежность к морской популяции определялась по характерному для морских особей фенотипу с большим количеством костных пластин на боковой поверхности тела. Фенотипически пресноводные особи собирались в озере Лобанешское на острове Великий, а также в озере Машинное на материке. Возраст популяций определялся как количество лет спустя момента начала опреснения, которое определялось исходя из высоты озёр относительно уровня моря. Возрасты популяций из Лобанешского и Машинного таким образом были определены равными ~600 и ~700 лет соответственно [75]. Данные популяции были использованы для детектирования маркерных ОНП по причине того, что они были одними из наиболее старых в области бассейна Белого моря. Тем самым, использование двух популяций старшего возраста даёт возможность поиска параллельных генетических изменений, которые, вероятно, содержат участки, вовлечённые в возникновение адаптации к пресной воде.



Рисунок 38. Карта расположения мест сбора образцов. Подробное описание образцов приведено в таблице 4.

Всего было получено более 18.000 маркерных ОНП, которые отличали морскую популяцию от пресноводных. Маркерные ОНП были распределены неравномерно вдоль генома, и образовывали ярко выраженные кластеры («острова дивергенции», ОД; [32,50]), для которых наблюдались повышенные плотности распределения маркерных ОНП. Всего было использовано два критерия для определения ОД. ОД, найденный по сильному (слабому) критерию, определялся как непрерывный участок, в котором в каждой 10 Кб рамке генома было как минимум 10 сильных (20 слабых) маркерных ОНП. После этого полученные участки на расстоянии <40 Кб объединялись, так как рекомбинация маловероятна на таких коротких расстояниях [98,101]. Такой алгоритм достаточно хорошо описывает наблюдаемые данные (рисунок 39), а также приводит к выделению меньшего числа более длинных островов по сравнению с критериями, использованными в другой работе [32]. Среди 6.107 маркерных ОНП, полученных при помощи сильного критерия, 5.801 (95%) было расположено внутри ОД. Всего при объединении сильного и слабого критериев было обнаружено 19 ОД, которые были расположены на 10 из 21 хромосом генома G. aculeatus (рисунок 39, таблица 5) и покрывали в целом 3.301.948 нуклеотидов последовательности или 0,74% от всего генома. Неравномерное распределение маркерных ОНП вдоль генома, вероятно, частично может объясняться неравномерностью рекомбинации [102]. Для каждого ОД частоты пресноводных аллелей могут подниматься за счёт одного отбираемого локуса вследствие эффекта генетического хитч-хайкинга [103] или же за счёт того, что сразу несколько отбираемых локусов находятся на достаточно близком расстоянии друг от друга, и отбор действует на всех них вместе [104].



расстояние вдоль генома, Мб

Рисунок 39. Распределение маркерных ОНП, различающих морскую и пресноводные популяции, в геноме *G. aculeatus*. На рисунке присутствуют все хромосомы, которые содержат хотя бы один ОД. Для каждой хромосомы горизонтальная ось показывает координаты расположения в геноме 10 Кб окон, а вертикальная ось показывает число маркеров внутри каждого из этих окон. Красный цвет используется для значений, определённых по «сильным» маркерам, а синий цвет для соответствующих значений, определённых по «слабым» маркерам. На данном рисунке более мелкими ромбами изображены значения <20 для слабых маркеров. Чёрными полосами сверху обозначены границы ОД, а зелёные полосы отмечают места известных инверсий на хромосомах I, XI и XXI [32].

ОД	Начало	Конец	# марк.	Также		# генов,	# нс	SE	SG				
			ОНП	были		перес. с	ОНП						
				найдены в		найдены в		найдены в		ОД			
				[63]	[32]								
I-1	21.487.998	21.960.119	4.186	+	+	27	100	0,247	0,212				
II-1*	14.874.366	14.898.826	73	+	+	0	0	0,115	0,146				
IV-1	12.803.780	12.881.296	285	+	+	7	19	0,188	0,094				
IV-2	13.930.002	13.959.331	168		+	2	1	0,218	0,107				
IV-3	19.811.922	19.914.666	209	+	+	6	0	0,212	0,163				
IV-4	23.954.634	23.981.981	48	+	+	3	1	0,227	0,1				
IV-5*	26.016.955	26.166.536	252	+	+	4	1	0,196	0,036				
V-1	2.482.209	2.501.295	65		+	2	3	0,255	0,239				
VII-1*	17.982.351	18.002.671	84	+	+	4	6	<0	0,273				
IX-1*	8.521.935	8.537.559	44		+	2	0	0,062	0,026				
IX-2*	8.901.816	8.910.115	20		+	2	0	0,108	0,023				
IX-3	9.208.158	9.227.809	46		+	1	0	0,198	0,131				
IX-4	10.334.101	10.353.801	114		+	1	0	0,141	0,101				
XI-1*	5.445.757	5.855.124	1.237	+	+	21	10	0,054	0,147				
XII-1	14.338.229	14.358.336	91	+	+	2	0	0,023	0,193				
XII-2	16.522.028	16.538.810	24		+	2	0	0,162					
XIX-1	2.449.903	2.581.858	277	+	+	5	11	0,178	0,131				
XIX-2	14.787.904	14.799.088	21	+	+	0	0	0,117	0,056				
XXI-1*	5.759.879	7.486.635	6.900	+	+	79	63	0,177	<0				
Всего			14.144	12	19	170	215						

Таблица 5. 19 ОД, отличающих морскую и пресноводные популяции. Идентификаторы ОД представлены в следующем формате '№ хромосомы-порядковый № ОД'. Звёздочкой отмечены ОД, которые были определены при помощи слабого критерия; все остальные ОД были определены при помощи сильного критерия. '+' отмечает ОД, которые были также найдены в работе [63] или в [32]; нс – число несинонимических замен,  $s_E$  – коэффициент отбора, способствующий закреплению пресноводного аллеля, рассчитанный для популяции озера Ершовского;  $s_G$  – коэффициент отбора, способствующий закреплению пресноводного аллеля, рассчитанный для популяции карьера Голубой.

В большинстве ОД количество слабых маркеров только немного превышало количество сильных маркеров, свидетельствуя о том, что пресноводные аллели в среднем достигли частоты 80% в обеих пресноводных популяциях. Двумя наиболее явными исключениями были острова IV-5 и XI-1 (пятый ОД 4-й хромосомы и первый ОД на 11-й хромосоме), которые, несмотря на то что были идентифицированы двумя критериями, содержали в ~20 раз больше слабых маркеров по сравнению с количеством маркеров, определённых по сильному критерию. Также пять ОД (ІІ-1, VII-1, IX-1, IX-2 и XXI-1) находились только при помощи слабого критерия, что свидетельствует о том, что частота пресноводного аллеля в них в одной или обеих популяциях была <0,8. В части таких ОД частота пресноводных аллелей не достигала порога вследствие недостаточно сильного отбора в одной или обеих популяциях. В ОД XXI-1 наблюдался быстрый рост частот пресноводных аллелей в молодых пресноводных популяциях, который затем замедлялся. В результате частота в молодых пресноводных популяциях не сильно отличалась от соответствующих частот в более старых пресноводных популяциях (рисунок 39). Наблюдаемый паттерн соотносится с предполагаемым действием балансирующего отбора, и возможно, на него могут влиять взаимодействия между многочисленными генами, которые находятся в пределах длинной инверсии внутри ОД XXI-1.

Внутри ОД частоты пресноводных аллелей сильно отличаются между морской и пресноводными популяциями. Количество различий может превышать 1% от всех нуклеотидных позиций для отдельных ОД. Так как большинство маркерных ОНП, вероятно, являются нейтральными, получается, что гаплотипы дивергировали ~10<sup>6</sup> поколений назад, принимая скорость мутирования равной 10<sup>-8</sup> на нуклеотид на поколение [88]. Такой высокий уровень дивергенции не соотносится с образованием аллелей за счёт *de novo* мутаций в каждой отдельной популяции. Вместо этого данное наблюдение соответствует повторяемой адаптации за счёт предсуществующей генетической изменчивости с использованием одних и тех же древних

аллелей в процессе основания независимых популяций. В пользу данного предположения были получены результаты в недавней работе [60].

Несмотря на то, что предковые морские и потомковые пресноводные аллели сосуществовали в популяции в течение долгого времени, они вероятно имели не так много возможностей быть разделёнными рекомбинацией от близлежащих нейтральных ОНП. Также стоит понимать, что рекомбинация, которая способна перемешивать морские и пресноводные гаплотипы, может это сделать только в случае, если ещё не произошла фиксация одного из гаплотипов в популяции. Таким образом, длина ОД может зависеть от силы отбора, скорости рекомбинации и количества поколений, которые гаплотип находился в популяции.

Общее нуклеотидное разнообразие ( $\pi$ ) у двух пресноводных образцов было снижено на 22% по сравнению с соответствующим значением в образцах морских популяций, что соответствует более низкой эффективной численности ( $N_e$ ) пресноводных популяций на настоящее время и/или момент основания популяций. Сильные и слабые ОД в морской популяции имели повышенное  $\pi$  (0,0049 и 0,0034) по сравнению со средним значением по геному (0,0020). В пресноводных популяциях  $\pi$  было понижено в сильных ОД (0,0012), но повышено в слабых ОД (0,0048) по сравнению со средним значением по геному (0,0026).

Нуклеотидное разнообразие внутри сильных ОД у пресноводных популяций немного ниже среднего значения, но не значительно, что может говорить об участии в адаптации мягких свипов [105] или являться результатом неоконченной фиксации, когда в популяции присутствуют морские и пресноводные гаплотипы. При адаптации за счёт мягких свипов сразу несколько гаплотипов поднимаются в частоте, и вследствие этого наблюдается не такое сильное снижение нуклеотидного разнообразия вокруг участка под отбором. Мягкие свипы более ожидаемы при адаптации за счёт предсуществующей генетической изменчивости, чем при адаптации за счёт *de novo* мутаций [106], и это также соотносится с транспортерной гипотезой [61]. Также ожидается, что нуклеотидное разнообразие будет повышено на границах острова по сравнению с центральной областью [107]. Данный эффект будет наблюдаться вследствие того, что при основании популяции несколько индивидуумов могут нести в себе пресноводный гаплотип, который в разных индивидуумах будет иметь различные границы.

17 из 19 найденных ОД перекрывались с границами 170 генов. Среди маркерных ОНП, находящихся внутри этих генов, 139 были несинонимическими и 146 были синонимическими (таблица 6). Внутри ОД отношение несинонимических к синонимическим маркерным ОНП было выше, чем то же отношение для немаркерных ОНП, которые были полиморфными внутри морской популяции (290 несинонимических и 528 синонимических; таблица 6). Данный факт

даёт основания предполагать, что большее количество позиций находится под положительным отбором среди маркерных ОНП, чем среди немаркерных ОНП [108]. Тем не менее, 7 из 17 ОД не содержат несинонимических маркерных ОНП и 2 ОД совсем не перекрываются ни с одним из известных генов. Предположительно гены, которые перекрываются с ОД, могут быть ответственны за фенотипические различия между морской и пресноводной формами колюшки. Среди таких генов широко известный ген EDA, ответственный за количество костных пластин на боковой поверхности тела рыбы ([59], ОД IV-1). Также с найденными ОД перекрываются гены, которые вероятно играют важную роль в процессе адаптации за счёт осморегуляции, иммунитета или морфологии: Na<sup>+</sup>/K<sup>+</sup>-АТФаза (ATP1A1 [109,110], ОД I-1), нейротрансмиттер и гормонсвязывающий ген (SULT4A1 [32], ОД IV-4), а также ген иммунного ответа к вирусной инфекции (NLRC5 [111], ОД XIX-1). Также в ОД наблюдались другие гены, которые имеют важные функциональные значения, например гены, вовлечённые в процессы метаболизма и поведения, (INHA [112], ОД I-1), ответственные за рост и развитие нервных клеток (SLITRK2 [113], ОД IV-5), адгезию и дифференцировку нервных клеток (CTNNA2 [114], ОД IX-3), обмен кальция и фосфата (STC2 [115], ОД IV-2), а также регулирование функций центральной и периферической нервной системы (HTR3A [116], ОД V-1). ОД, которые не содержат генов с несинонимическими заменами, предположительно могут играть роль в адаптации, выполняя регуляторные функции близлежащих генов, что соотносится с результатами, полученными в предыдущей работе [32].

	Маркерные ОН	Π		ОНП морской популяции			
	Несинонимич.	Синонимич.	Отношение	Несинонимич.	Синонимич.	Отношение	
Внутри	139	146	0,95	290	528	0,55	
од							
Вне ОД	17	4	4,25	28.616	41.902	0,68	

Таблица 6. В таблице представлены числа несинонимических и синонимических замен, расположенных внутри и вне ОД, для маркерных ОНП и для немаркерных ОНП морской популяции.

Только 5% маркерных ОНП были расположены за пределами ОД. Всего таких было 306 маркерных ОНП, которые располагались на 19 из 21 хромосом. 21 из 306 таких маркерных ОНП были расположены внутри генов, и из них 17 были несинонимическими заменами между морской и пресноводной популяциями, в то время как только 4 были синонимическими. В данном случае,

как и в случае с маркерными ОНП внутри ОД, отношение было больше по сравнению с соответствующим отношением для полиморфизма морской популяции (для неё значения несинонимических и синонимических замен составляли соответственно 28.616 и 41.902; таблица 6). Полученный результат соответствует тому, что положительный отбор действует не только внутри островов, но и за их пределами. Всего было найдено 12 генов, которые содержали несинонимические маркерные ОНП за пределами ОД, и они, вероятно, могут участвовать в адаптации колюшки. К примеру, ген GCNT3, который играет важную роль в биосинтезе муцина, использующегося в качестве клейкого вещества при постройке гнёзд [117], содержал 4 маркерных ОНП, все из которых были несинонимическими. Другой ген, играющий роль в размножении, *MUC* с хромосомы II содержал 2 несинонимические замены. Два несинонимических маркерных ОНП за пределами ОД располагались на расстоянии 300 нуклеотидов внутри гена *EPX* на хромосоме XIII. *EPX* является геном, играющим роль в активности эозинофилов, которые осуществляют лизис паразитов [118]. Ещё один несинонимический ОНП был расположен в пределах гена *INSR* на хромосоме III; который играет важную роль в росте и раннем развитии, а также в развитии нервной системы [119,120].

Часть маркерных ОНП за пределами ОД в некоторых случаях образовывала кластеры, которые не проходили порог на минимальное количество маркеров для определения ОД. Предположительно, частота некоторых маркерных ОНП может подниматься за счёт генетического хитч-хайкинга с близлежащими ОД, хотя частота пресноводных аллелей маркерных ОНП лишь слабо скоррелирована с расстоянием от ОД (рисунок 40). С другой стороны, вероятно, некоторые из этих маркерных ОНП образовались недавно и поэтому ещё не накопили вокруг себя нейтральные маркеры за счёт хитч-хайкинга, и в этом случае вне ОД ожидается большее количество ОНП под действием отбора. Однако каждый отдельно взятый маркерный ОНП вне острова с большей вероятностью является результатом дрейфа или ошибкой, чем отдельный ОД, который состоит из большого количества маркерных ОНП. Кроме того, вне ОД наблюдается повышение частот пресноводных аллелей для маркерных ОНП, и это возрастание не такое сильное по сравнению с ростом частот маркерных ОНП внутри ОД, что может быть результатом большего количества нейтральных аллелей или более слабого отбора за пределами ОД (рисунок 40).



Рисунок 40. Средние частоты пресноводных аллелей в маркерных ОНП, полученные при помощи сильного критерия и находящиеся за пределами ОД. Показана зависимость частот от расстояния до ОД. Горизонтальная ось показывает расстояние от маркерного ОНП до ближайшего ОД в нуклеотидах (логарифмическая шкала). Вертикальная ось показывает частоту пресноводного аллеля в маркерном ОНП.

#### 2.2.2. Динамика адаптации G. aculeatus к пресноводной среде обитания

Помимо образцов из озёр Машинное и Лобанешское также были секвенированы образцы *G. aculeatus* из четырёх более молодых популяций: карьеров Голубой и Малыш, озёр Марцы и Ершовское (таблица 4), все из которых располагаются вблизи от Белого моря. Карьерные популяции были основаны В. Зюгановым в 1978 г. в заброшенных изолированных карьерах, в которых до заселения не было рыбы этого вида. Таким образом, возраст этих популяций составляет 34 года на момент сбора образцов. Карьер Голубой (площадью ~70.000 м<sup>2</sup> и вместимостью ~1.000 рыб) был изначально заселён 20 морскими (10 самок и 10 самцов) и 20 пресноводными (10 самок и 10 самцов) особями. Карьер Малыш (площадью ~75 м<sup>2</sup> и вместимостью ~100 рыб, с очень малым количеством мест пригодных для постройки гнёзд) был заселён 1 морской самкой и 1 пресноводным самцом [71]. Морские особи были взяты из популяции Белого моря, а пресноводные особи были из популяции озера Машинное. Озёра Марцы и Ершовское образовались вследствие изоляции морских заливов в результате
постепенного поднятия земной коры над уровнем моря со скоростью  $\sim$ 3,8 мм за год [75]. Возраст пресноводной популяции озера Марцы составляет  $\sim$ 250 лет вследствие наблюдения, что поверхность озера на данный момент находится на возвышении  $\sim$ 1 м над уровнем моря. Озеро Ершовское, которое на данный момент находится на возвышении  $\sim$ 13 см над уровнем моря, ещё в 1978 году представляло собой озеро, населённое только проходными особями с фенотипом, характерным для морской популяции [72]. В данное время в нём находится достаточно большая популяция резидентных особей, имеющих характерную морфологию и специфические для данного фенотипа паразиты (вид *Schistocephalus solidus* является одним из доминирующих видов паразитов для резидентных особей, а нематоды характерны для морских индивидуумов [71]). Вследствие всего вышесказанного мы предполагаем возраст популяции озера Ершовское равным 34 годам. По причине малых возрастов, вероятно, что во всех вышеперечисленных молодых пресноводных популяциях процесс адаптации все ещё продолжается.

Частоты пресноводных аллелей в маркерных ОНП внутри островов во всех четырёх молодых пресноводных популяциях значительно увеличены относительно соответствующих в морской популяции (рисунки 41-42, см. таблицу 7 для данных о частотах аллелей для всех популяций). Для пяти ОД была проведена валидация частот пресноводных аллелей путём генотипирования особей из каждой популяции аллель-специфическими праймерами (таблицы 7-8). Наблюдаемая динамика роста частоты пресноводного аллеля с увеличением возраста популяции предполагает, что в каждой из них отбор благоприятствует одним и тем же аллелям. Первоначальные частоты пресноводных аллелей в популяциях озёр Ершовское и Марцы предположительно были равными наблюдаемым в морской популяции (~0,1, таблица 5). В двух искусственных популяциях, населённых равными количествами морских и пресноводных индивидуумов, изначальная средняя частота пресноводного аллеля составляла 0,5. Измеренные средние частоты пресноводных аллелей искусственных популяций составляли 0,56 для карьера Малыш и 0,73 для карьера Голубой (рисунок 42). В карьере Малыш частоты пресноводных аллелей имели меньшее среднее и большую дисперсию по сравнению с соответствующими для карьера Голубой, что вероятно может являться результатом более низкой эффективной численности карьера Малыш и вследствие этого более сильным эффектом дрейфа в этой популяции. Частоты пресноводных аллелей в озере Марцы в среднем выше, чем в озере Ершовское, вероятно вследствие того, что последняя популяция имеет меньший возраст.



Рисунок 41. Средние частоты пресноводных аллелей в маркерных ОНП внутри ОД в популяциях разных возрастов. (А) Природные популяции. Горизонтальная ось показывает примерный возраст популяций, слева направо идут: морская популяция (~0 лет), молодые пресноводные популяции (~34 и ~250 лет) и более старые пресноводные популяции (~600 и ~700 лет). Усы соответствуют стандартным отклонениям. (Б) Искусственные популяции. Для каждого ОД первоначальная частота пресноводных аллелей считается равной 50% (чёрная линия), а частоты в настоящем времени показаны: карьер Малыш (слева) и карьер Голубой (справа), популяции которых были основаны в 1978 г. Усы, прямоугольники и прямые внутри них соответствуют 5-й и 95-й перцентилям, стандартному отклонению и медиане соответственно.

Красный цвет – ОД, определённые по сильному критерию; синий – ОД, определённые по слабому критерию.

ОД	Нильма	Ершовск.	Ершовск.	Марцы	Голубой	Малыш	Лобанешс.	Машин.
		(проходн.)	(резиден.)					
I-1	0,07±0,05	0,06±0,06	0,54±0,09	0,77±0,15	0,87±0,11	0,56±0,09	1±0,01	1±0,01
II-1*	0,1±0,06	0,05±0,06	0,23±0,04	0,27±0,11	0,78±0,06	0,59±0,05	0,71±0,08	0,66±0,1
IV-1	0,12±0,06	0,07±0,07	0,38±0,09	0,67±0,12	0,69±0,09	0,5±0,09	1±0,02	0,99±0,03
IV-2	0,12±0,05	0,1±0,06	0,46±0,09	0,67±0,11	0,71±0,06	0,52±0,08	1±0,01	1±0
IV-3	0,14±0,05	0,12±0,06	0,44±0,1	0,68±0,12	0,8±0,12	0,61±0,1	1±0,02	1±0,02
IV-4	0,15±0,04	0,08±0,07	0,48±0,08	0,73±0,11	0,7±0,07	0,67±0,05	1±0	0,99±0,05
IV-5*	0,14±0,04	0,11±0,06	0,4±0,1	0,71±0,13	0,57±0,09	0,67±0,07	1±0,03	0,69±0,12
V-1	0,11±0,05	0,13±0,06	0,56±0,1	0,73±0,13	0,89±0,11	1±0	1±0,02	1±0
VII-1*	0,08±0,04	0,05±0,06	0,05±0,05	0,23±0,11	0,92±0,11	0,83±0,08	0,58±0,07	1±0
IX-1*	0,11±0,05	0,04±0,06	0,16±0,07	0,43±0,11	0,55±0,09	0,69±0,06	0,55±0,04	0,79±0,15
IX-2*	0,13±0,04	0,08±0,07	0,22±0,09	0,46±0,12	0,55±0,09	0,16±0,28	0,63±0,15	0,72±0,2
IX-3	0,14±0,06	0,05±0,06	0,41±0,09	0,66±0,09	0,75±0,1	0,67±0,06	0,97±0,05	1±0
IX-4	0,12±0,05	0,07±0,06	0,28±0,08	0,73±0,09	0,7±0,07	0,9±0,08	1±0,01	1±0,01
XI-1*	0,05±0,05	0,03±0,05	0,15±0,09	0,48±0,12	0,78±0,1	0,1±0,08	0,73±0,11	0,79±0,14
XII-1	0,15±0,04	0,1±0,06	0,12±0,06	0,41±0,11	0,84±0,11	0,83±0,08	1±0,02	1±0
XII-2	0,08±0,06	0,05±0,07	0,32±0,07	0,41±0,1	0,98±0,03	0,29±0,06	0,94±0,08	0,99±0,02
XIX-1	0,13±0,05	0,07±0,06	0,36±0,1	0,92±0,13	0,75±0,12	0,68±0,09	1±0,02	1±0,01
XIX-2	0,09±0,06	0,05±0,06	0,24±0,07	0,49±0,18	0,61±0,12	0,37±0,13	0,99±0,04	0,98±0,05
XXI-1*	0,01±0,02	0,03±0,05	0,36±0,09	0,59±0,13	0,47±0,1	0,01±0,05	0,69±0,11	0,59±0,08

Таблица 7. Средние частоты пресноводных аллелей в каждой популяции для каждого из 19 ОД, среднее ± стандартное отклонение. Звёздочки обозначают ОД, определённые по слабому критерию.

ОД	Нильма		Ершовское		Ершовское		Марцы	
			(проходные)		(резидентные)			
	ПЦР	Иллюмина	ПЦР	Иллюмина	ПЦР	Иллюмина	ПЦР	Иллюмина
I-1	0,125	0,07±0,05	0	0,06±0,06	0,500	0,54±0,09	0,750	0,77±0,15
IV-1	0,125	0,12±0,06	0,125	0,07±0,07	0,417	0,38±0,09	0,700	0,67±0,12
IV-3	0,156	0,14±0,05	0,15	0,12±0,06	0,458	0,44±0,1	0,750	0,68±0,12
IV-4	0,156	0,15±0,04	0,2	0,08±0,07	0,458	0,48±0,08	0,650	0,73±0,11
IX-4	0,156	0,12±0,05	0,05	0,07±0,06	0,208	0,28±0,08	0,700	0,73±0,09
XIX-1a	0,094	0,13±0,05	-	0,07±0,06	0,375	0,36±0,1	0,750	0,92±0,13
XIX-1b	0,094	0,13±0,05	0,05	0,07±0,06	0,250	0,36±0,1	0,850	0,92±0,13
XXI-1*	0,000	0,01±0,02	-	0,03±0,05	0,333	0, <del>36±0,09</del>	0,550	0,59±0,13

ОД	Голубой		Малыш		Лобанешское		Машинное	
	ПЦР	Иллюмина	ПЦР	Иллюмина	ПЦР	Иллюмина	ПЦР	Иллюмина
I-1	0,850	0,87±0,11	0,525	0,56±0,09	1,000	1±0,01	1,000	1±0,01
IV-1	0,700	0,69±0,09	0,525	0,5±0,09	0,938	1±0,02	0,900	0,99±0,03
IV-3	0,875	0,8±0,12	0,650	0,61±0,1	1,000	1±0,02	1,000	1±0,02
IV-4	0,700	0,7±0,07	0,675	0,67±0,05	1,000	1±0	0,900	0,99±0,05
IX-4	0,750	0,7±0,07	0,700	0,9±0,08	1,000	1±0,01	1,000	1±0,01
XIX-1a	0,775	0,75±0,12	0,650	0,68±0,09	1,000	1±0,02	1,000	1±0,01
XIX-1b	0,775	0,75±0,12	0,675	0, <del>68±0,09</del>	1,000	1±0,02	1,000	1±0,01
XXI-1*	0,350	0,47±0,1	0,000	0,01±0,05	0,688	0, <del>6</del> 9±0,11	0,450	0,59±0,08

Таблица 8. Сравнение частоты пресноводных аллелей в ОД, определённых при помощи аллель-специфических праймеров по сравнению с рассчитанными по частотам маркерных ОНП.

Также оказалось, что частоты пресноводных аллелей в ОНП были повышены и за пределами ОД. Данное наблюдение было сделано для двух природных молодых популяций для всех хромосом, кроме хромосом XIV и XVI, каждая из которых имела только по одному маркерному ОНП (рисунок 42A). То же наблюдение было сделано для большинства хромосом карьера Голубой (рисунок 42Б). Тем не менее, в среднем частоты маркерных ОНП внутри ОД имели более высокие значения по сравнению с маркерными ОНП за пределами ОД во всех пресноводных популяциях (рисунок 43). Увеличение частот пресноводных аллелей также наблюдалось в случае, когда только одна пара популяций использовалась для поиска маркерных ОНП (в качестве морской использовалась популяция Нильмы и в качестве пресноводной – популяция из Машинного; рисунок 44).



Рисунок 42. Средние частоты пресноводных аллелей в маркерных ОНП, полученные при помощи сильного критерия и находящиеся за пределами ОД в пресноводных популяциях различных

возрастов. (А) Природные популяции. Горизонтальная ось показывает примерные возрасты популяций, начиная слева-направо с морской популяции (~0 лет), затем идут наиболее молодые популяции (~34 и ~250 лет) и потом следуют более старые пресноводные популяции (~600 и ~700 лет). Усы соответствуют стандартным отклонениям. (Б) Природные популяции. Для каждого ОД изначальная частота пресноводных аллелей (чёрная линия, 50%) и соответствующая частота на момент сбора образцов показаны для двух искусственных популяций: карьеров Малыш (слева) и Голубой (справа), каждый из которых был основан в 1978 году. Усы, прямоугольники и прямые внутри них соответствуют 5-й и 95-й перцентилям, стандартному отклонению и медиане соответственно.



Рисунок 43. Сравнение средних частот пресноводных аллелей в маркерных ОНП внутри и снаружи ОД в молодых пресноводных популяциях. Каждые два прямоугольника соответствуют популяциям слева-направо: пресноводные из озера Ершовское, озера Марцы, карьера Малыш, карьера Голубой. Р-значения, определённые при помощи двустороннего критерия Манна-Уитни, <0,001 отмечены \*\*\*. Усы, прямоугольники и прямые внутри них соответствуют 5-й и 95-й перцентилям, стандартному отклонению и медиане соответственно. Красный цвет – маркерные ОНП, определённые по сильному критерию, внутри ОД; фиолетовый – маркерные ОНП, определённые по сильному критерию, за пределами ОД.



Рисунок 44. Сравнение средних частот пресноводных аллелей в маркерных ОНП внутри и снаружи ОД, определённых при участии в сравнении только одной морской и одной пресноводной популяций. Показаны данные для пресноводных популяций различных возрастов. Каждые два прямоугольника (красный и фиолетовый) соответствуют четырём молодым пресноводным популяциям. Усы, прямоугольники и прямые внутри них соответствуют 5-й и 95-й перцентилям, стандартному отклонению и медиане соответственно. Красный цвет соответствует маркерным ОНП (по сильному критерию), расположенные внутри ОД; фиолетовым обозначены маркерные ОНП (по сильному критерию), расположенные снаружи ОД.

#### 2.2.3. Сила отбора, способствующего адаптации к пресноводной среде обитания

При знании изменений частот пресноводных аллелей можно рассчитать силу отбора, который влияет на них [100,121]. Вследствие того, что возраст озера Марцы известен только приблизительно, было решено не использовать эту популяцию для расчёта силы отбора. Также для расчётов не была использована популяция карьера Малыш, так как она имела изначально очень низкую численность, и в таком случае ожидается сильный эффект действия генетического дрейфа. Тем самым, расчёты проводились только для молодых пресноводных популяций озера Ершовское и карьера Голубой. При оценке коэффициента отбора в качестве значения длины поколения *G. aculeatus* было взято 2 года ([53] и см. методы), а в качестве коэффициента доминирования был выбран h=0,5, что согласуется с результатами недавних исследований для множества морфологических признаков *G. aculeatus* [122].

Оценки коэффициентов отбора были рассчитаны для ОД, а не для отдельных маркерных ОНП, так как оценка частоты в индивидуальных ОНП будет смещена с большей вероятностью по сравнению с оценкой для ОД. Коэффициент отбора *s* рассчитывался на основе средней частоты пресноводных аллелей внутри ОД. Стоит понимать, что частоты маркерных ОНП не являются независимыми внутри ОД, и это позволяет более точно рассчитать среднюю частоту пресноводных аллелей внутри ОД и соответственно получить более достоверную оценку коэффициента отбора.

В целом, полученные значения для *s* были достаточно высокими: из 19 ОД 15 в популяции озера Ершовское и 12 в популяции карьер Голубой имели s>0,1 (таблица 6, рисунок 45). Коэффициенты отбора, оцененные для каждого ОД, не имели значимой корреляции между двумя пресноводными популяциями (рисунок 45; p=0,30, P=0,27). Данный результат ожидаем, поскольку мы рассматриваем лишь небольшое количество ОД, и ошибки оценок коэффициентов отбора для отдельных ОД достаточно высоки. Тем не менее, средние значения для *s* популяций озера Ершовское (s=0,16) достаточно схожи с s для популяции из карьера Голубой (s=0,13). В частности, ОД IV-1, в котором расположен ген EDA [59] имеет коэффициент отбора *s*=0,19 в популяции озера Ершовское и s=0.09 в популяции карьера Голубой. Данные высокие значения s согласуются с ранее полученными данными о том, что изменения фенотипа по признаку, контролируемому геном EDA, наблюдаются уже на очень коротких временных промежутках при изменении условий среды обитания [42]. Наибольшее изменение средней частоты пресноводного аллеля в популяции озера Ершовское было обнаружено для ОД V-1, в котором соответствующая частота возросла с 0,1 до 0,56, что соответствует коэффициенту отбора s=0,255. В данном ОД наблюдается несколько несинонимических замен в гене *HTR3A*, субъединице серотонинового рецептора с широким кругом физиологических функций. С немного меньшим изменением частоты наблюдался ОД I-1 (*s*<sub>E</sub>=0,247,  $s_G=0,212$ ), который находится в инверсии длиной 470 Кб, содержащей 27 генов, включая ген ATP1A1, кодирующий белок Na<sup>+</sup>/K<sup>+</sup> АТФазу. Маркерные ОНП внутри этого гена включают 6 несинонимических замен, которые могут быть частично ответственны за различия в осморегуляции между морской и пресноводной популяциями [123]. Среди ОД популяции карьера Голубой наиболее сильное изменение частоты пресноводного аллеля наблюдалось для ОД XII-2. В нём частота почти достигала 1, и при таких значениях *s* достаточно трудно рассчитать точно. Данный ОД перекрывается с участком, находящимся вблизи от 5'-конца гена OVGP1, эстроген зависимого гена женской репродуктивной системы. Данный ген участвует в процессе размножения [124], и тем самым является вероятной мишенью для действия положительного отбора.



Рисунок 45. Коэффициенты отбора для ОД, определённые для популяций из озера Ершовское и карьера Голубой. Красным обозначены *s* в ОД, определённых по сильному критерию, синим – *s* в ОД, определённых по слабому критерию. Три ОД не обозначены на рисунке: VII-1 и XXI-1, для которых s<0 в одной из популяций, и XII-2, для которой *s* в популяции карьера Голубой не может быть рассчитан относительно достоверно из-за частоты пресноводного аллеля, близкой к фиксации (таблица 5). Показаны *s*, полученные для оценок средних частот пресноводных аллелей в ОД с учётом стандартных отклонений.

В предыдущей работе рассчитывался коэффициент отбора при адаптации к пресной воде колюшки для гена *EDA* [42] по результатам изменения частот в краткосрочном эволюционном эксперименте. Полученные значения в предыдущей работе (s~0,5) были несколько выше по сравнению с полученными в данном исследовании (s=0,19 и 0,09). Подобные расчёты коэффициентов отбора во время процесса адаптации производились также и у других организмов, например, у вида *Biston betularia*, берёзовой пяденицы, при адаптации против истребления хищниками (s=0,05–0,16 [125]). Другим примером является адаптация к усваиванию лактозы у человека (s = 0,014–0,19 [126]). Также оценки s, полученные в данной работе, превышают те, что были рассчитаны для промотора гена *PDYN* (s<0,01 [127]) и генов, вовлечённых в пигментацию у человека (s=0,02–0,10, [128]).

Для некоторых ОД наблюдаются достаточно высокие коэффициенты отбора. Вероятно, отбор действует на низкочастотные аллели, которые уже адаптированы к пресноводной среде обитания. В таком случае наличие пресноводных аллелей на момент основания пресноводной популяции позволяет ей быстро адаптироваться к новой среде с отличающимися солёностью, паразитами и другими условиями. Предположительно, данный тип адаптации может быть распространён за пределами вида *G. aculeatus*. Несколькими из ряда других примеров являются быстрая адаптация видов родов *Eubosmina* в европейских озёрах [129], *Labeobarbus* в озёрах и реках Эфиопии [130] и *Salvelinus* [131].

#### 2.3. Выводы

Трёхиглая колюшка является широко используемой моделью для изучения адаптации. Исследование множества независимых пресноводных популяций позволяет лучше понять, как происходит адаптация к похожей среде обитания несколько раз независимо, а также изучить особенности адаптации за счёт предсуществующей генетической изменчивости в условиях действия сильного отбора.

В результате исследования архитектуры адаптации трёхиглой колюшки к пресноводной среде обитания были найдены участки, в которых повышена плотность маркерных ОНП, различающих морские и пресноводные популяции Белого моря, острова дивергенции. При изучении частот пресноводных аллелей в молодых популяциях колюшки были выявлены скорости роста частот ОД и произведены оценки коэффициентов отбора для них. Данные оценки получены по большому количеству поколений, что повышает их точность. Также были оценены соотношения частот несинонимических и синонимических замен и показано, что большие соотношения наблюдаются снаружи ОД. Полученные результаты вносят вклад в понимание процесса адаптации, а также начальной стадии процесса видообразования, разработанные методы и система анализа могут быть использованы при изучении других систем, в которых происходит адаптация с использованием предсуществующей генетической изменчивости.

## Список публикаций по теме диссертации

1. Terekhanova N.V., Logacheva M.D., Penin A.A., Neretina T.V., Barmintseva A.E., Bazykin G.A., Kondrashov A.S., Mugue N.S. Fast evolution from precast bricks: genomics of young freshwater populations of threespine stickleback *Gasterosteus aculeatus* // PLoS Genet 2014. Vol. 10, no. 10, P. e1004696.

2. Terekhanova N.V., Seplyarskiy V.B., Soldatov R.A., Bazykin G.A. Evolution of local mutation rate and its determinants // Mol Biol Evol 2017. Vol. 34, no. 5, P. 1100–1109.

### Тезисы конференций

 Terekhanova N.V., Mugue N.S., Bazykin G.A., Kondrashov A.S. Population genomics of threespine stickleback after 30-year evolution experiment // "Society for Molecular Biology & Evolution" (SMBE 2012), Дублин, Ирландия, 2012.

2. Тереханова Н.В., Базыкин Г.А., Кондрашов А.С., Мюге Н.С. Genetic variation in the genomes of the threespine stickleback *Gasterosteus aculeatus* // "Информационные технологии и системы" (ИТИС 2012), Петрозаводск, Россия, 2012.

3. Terekhanova N.V., Bazykin G.A., Kondrashov A.S., Mugue N.S. Patterns of genomic divergence during adaptation to fresh water in the threespine stickleback *Gasterosteus aculeatus* // "Moscow Conference on Computational Molecular Biology" (MCCMB 2013), Москва, Россия 2013.

4. Terekhanova N.V., Bazykin G.A., Kondrashov A.S., Mugue N.S. Genome-wide patterns of divergence during adaptation to fresh water in threespine stickleback *Gasterosteus aculeatus* // "Информационные технологии и системы" (ИТИС 2013), Калининград, Россия, 2013.

5. Terekhanova N.V., Bazykin G.A., Kondrashov A.S., Mugue N.S. Fast evolution in young freshwater populations of *G. aculeatus* // "Society for Molecular Biology & Evolution" (SMBE 2014), Сан-Хуан, Пуэрто-Рико, 2014. 6. Terekhanova N.V., Bazykin G.A., Soldatov R.A., Seplyarskiy V.B. Local variation of the mutation rate across the primate phylogeny // "Moscow Conference on Computational Molecular Biology" (MCCMB 2015), Москва, Россия, 2015.

7. Terekhanova N.V., Bazykin G.A., Soldatov R.A., Seplyarskiy V.B. Evolution of local mutation rate and its determinants // "Информационные технологии и системы" (ИТИС 2015), Сочи, Россия, 2015.

# Благодарности

Я благодарна своему научному руководителю Базыкину Г.А., соавторам статей Барминцевой А.Е., Кондрашову А.С., Логачёвой М.Д., Мюге Н.С., Неретиной Т.В., Пенину А.А., Сеплярскому В.Б., Солдатову Р.А. и всем сотрудникам моей лаборатории за обсуждения и ценные советы.

# Список сокращений

- Кб килобаза, 1.000 нуклеотидов
- ЛСМ локальная скорость мутирования
- Мб мегабаза, 1.000.000 нуклеотидов
- ОД остров дивергенции
- ОНП однонуклеотидный полиморфизм

## Список литературы

- Feder J.L. et al. Establishment of new mutations under divergence and genome hitchhiking // Philosophical Transactions of the Royal Society B: Biological Sciences. 2012. Vol. 367, № 1587. P. 461–474.
- Flaxman S.M., Feder J.L., Nosil P. Genetic hitchhiking and the dynamic buildup of genomic divergence during speciation with gene flow // Evolution. 2013. Vol. 67, № 9. P. 2577–2591.
- Turner T.L., Hahn M.W., Nuzhdin S.V. Genomic islands of speciation in Anopheles gambiae // PLOS Biology. 2005. Vol. 3, № 9. P. e285.
- 4. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome // Nature. 2012. Vol. 489. P. 57.
- Bernstein B.E. et al. The NIH Roadmap Epigenomics Mapping Consortium // Nature biotechnology. 2010. Vol. 28, № 10. P. 1045–1048.
- 6. Schuster-Böckler B., Lehner B. Chromatin organization is a major influence on regional mutation rates in human cancer cells // Nature. 2012. Vol. 488. P. 504.
- Tyekucheva S. et al. Human-macaque comparisons illuminate variation in neutral substitution rates // Genome Biology. 2008. Vol. 9, № 4. P. R76.
- Karolchik D. et al. The UCSC Genome Browser database: 2014 update // Nucleic Acids Research.
   2014. Vol. 42, № Database issue. P. D764–D770.
- Lawrence M.S. et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes // Nature. 2013. Vol. 499. P. 214.
- Alexandrov L.B. et al. Signatures of mutational processes in human cancer // Nature. 2013. Vol. 500.
   P. 415.
- Nusbaum C. et al. DNA sequence and analysis of human chromosome 8 // Nature. 2006. Vol. 439. P. 331.
- Hodgkinson A., Eyre-Walker A. Variation in the mutation rate across mammalian genomes // Nature Reviews Genetics. 2011. Vol. 12. P. 756.
- Makova K.D., Hardison R.C. The effects of chromatin organization on variation in mutation rates in the genome // Nature Reviews Genetics. 2015. Vol. 16. P. 213.
- Stamatoyannopoulos J.A. et al. Human mutation rate associated with DNA replication timing // Nature Genetics. 2009. Vol. 41. P. 393.
- 15. Альбертс Б. et al. Молекулярная биология клетки. 5th ed. Ижевск: Институт компьютерных исследований, 2012.

- 16. Arbeithuber B. et al. Crossovers are associated with mutation and biased gene conversion at recombination hotspots // Proc Natl Acad Sci USA. 2015. Vol. 112, № 7. P. 2109.
- 17. Stevison L.S. et al. The time scale of recombination rate evolution in great apes // Molecular Biology and Evolution. 2016. Vol. 33, № 4. P. 928–945.
- Pratto F. et al. Recombination initiation maps of individual human genomes // Science. 2014. Vol. 346, № 6211.
- 19. Cole F., Keeney S., Jasin M. Comprehensive, fine-scale dissection of homologous recombination outcomes at a hot spot in mouse meiosis // Molecular Cell. 2010. Vol. 39, № 5. P. 700–710.
- Vermeulen W., Fousteri M. Mammalian transcription-coupled excision repair // Cold Spring Harbor Perspectives in Biology. 2013. Vol. 5, № 8. P. a012625.
- 21. Haradhvala N.J. et al. Mutational strand asymmetries in cancer genomes reveal mechanisms of DNA damage and repair // Cell. 2016. Vol. 164, № 3. P. 538–549.
- 22. Yazdi P.G. et al. Increasing nucleosome occupancy is correlated with an increasing mutation rate so long as DNA repair machinery is intact // PLOS ONE. 2015. Vol. 10, № 8. P. e0136574.
- 23. Imakaev M. et al. Iterative correction of Hi-C data reveals hallmarks of chromosome organization // Nature methods. 2012. Vol. 9, № 10. P. 999–1003.
- 24. Yu W., He B., Tan K. Identifying topologically associating domains and subdomains by Gaussian Mixture model And Proportion test // Nature Communications. 2017. Vol. 8. P. 535.
- 25. Sunyaev S. et al. Impact of selection, mutation rate and genetic drift on human genetic variation // Human Molecular Genetics. 2003. Vol. 12, № 24. P. 3325–3330.
- 26. Woo Y.H., Li W.-H. DNA replication timing and selection shape the landscape of nucleotide variation in cancer genomes // Nature Communications. 2012. Vol. 3. P. 1004.
- 27. Duret L., Arndt P.F. The impact of recombination on nucleotide substitutions in the human genome // PLOS Genetics. 2008. Vol. 4, № 5. P. e1000071.
- Thurman R.E. et al. The accessible chromatin landscape of the human genome // Nature. 2012. Vol. 489. P. 75.
- Kuruppumullage Don P. et al. Segmenting the human genome based on states of neutral genetic divergence // Proceedings of the National Academy of Sciences of the United States of America. 2013. Vol. 110, № 36. P. 14699–14704.
- 30. Imamura H., Karro J.E., Chuang J.H. Weak preservation of local neutral substitution rates across mammalian genomes // BMC Evolutionary Biology. 2009. Vol. 9, № 1. P. 89.

- Scally A. et al. Insights into hominid evolution from the gorilla genome sequence // Nature. 2012. Vol. 483. P. 169.
- 32. Jones F.C. et al. The genomic basis of adaptive evolution in threespine sticklebacks // Nature. 2012.
   Vol. 484, № 7392. P. 55–61.
- 33. Via S. Divergence hitchhiking and the spread of genomic isolation during ecological speciation-withgene-flow // Philosophical Transactions of the Royal Society B: Biological Sciences. 2012. Vol. 367, № 1587. P. 451–460.
- 34. Riesch R. et al. Transitions between phases of genomic differentiation during stick-insect speciation // Nature Ecology & Evolution. 2017. Vol. 1. P. 0082.
- 35. Hofer T., Foll M., Excoffier L. Evolutionary forces shaping genomic islands of population differentiation in humans // BMC Genomics. 2012. Vol. 13, № 1. P. 107.
- 36. Darwin C. On the origin of species by means of natural selection, or, the preservation of favoured races in the struggle for life. 1859.
- 37. Roux C. et al. Shedding light on the grey zone of speciation along a continuum of genomic divergence // PLoS Biology / ed. Moritz C. 2016. Vol. 14, № 12. P. e2000234.
- 38. Nadeau N.J. et al. Genomic islands of divergence in hybridizing Heliconius butterflies identified by large-scale targeted sequencing // Philosophical Transactions of the Royal Society B: Biological Sciences. 2012. Vol. 367, № 1587. P. 343–353.
- 39. Yeaman S., Aeschbacher S., Bürger R. The evolution of genomic islands by increased establishment probability of linked alleles // Mol Ecol. 2016. Vol. 25, № 11. P. 2542–2558.
- 40. Yeaman S., Whitlock M.C. The genetic architecture of adaptation under migretion-selection balance // Evolution. 2011. Vol. 65, № 7. P. 1897–1911.
- 41. Nosil P. Ecological speciation. Oxford University Press, 2012.
- 42. Barrett R.D.H., Rogers S.M., Schluter D. Natural Selection on a Major Armor Gene in Threespine Stickleback // Science. 2008. Vol. 322, № 5899. P. 255.
- 43. Soria-Carrasco V. et al. Stick insect genomes reveal natural selection's role in parallel speciation //
   Science. 2014. Vol. 344, № 6185. P. 738.
- 44. Grant P.R., Grant B.R. Unpredictable evolution in a 30-year study of darwin's finches // Science.
  2002. Vol. 296, № 5568. P. 707.
- 45. Tenaillon O. et al. Tempo and mode of genome evolution in a 50,000-generation experiment // Nature.2016. Vol. 536. P. 165.

- 46. Good B.H. et al. The dynamics of molecular evolution over 60,000 generations // Nature. 2017. Vol. 551. P. 45.
- 47. Kolbe J.J. et al. Founder effects persist despite adaptive differentiation: a field experiment with lizards
  // Science. 2012. Vol. 335, № 6072. P. 1086.
- 48. Rebeiz M. et al. Stepwise modification of a modular enhancer underlies adaptation in a Drosophila population // Science. 2009. Vol. 326, № 5960. P. 1663.
- 49. Elmer K.R. et al. Rapid evolution and selection inferred from the transcriptomes of sympatric crater lake cichlid fishes // Molecular Ecology. 2010. Vol. 19, № s1. P. 197–211.
- 50. Ellegren H. et al. The genomic landscape of species divergence in Ficedula flycatchers // Nature. 2012. Vol. 491, № 7426. P. 756–760.
- Liu S. et al. Population genomics reveal recent speciation and rapid evolutionary adaptation in polar bears // Cell. 2014. Vol. 157, № 4. P. 785–794.
- Conte G.L. et al. The probability of genetic parallelism and convergence in natural populations // Proc Biol Sci. 2012. Vol. 279, № 1749. P. 5039.
- 53. Bell M.A., Aguirre W.E., Buck N.J. Twelwe years of contemporary armor evolution in a threespine stickleback population // Evolution. 2007. Vol. 58, № 4. P. 814–824.
- 54. Bassham S. et al. Repeated selection of alternatively adapted haplotypes creates sweeping genomic remodeling in stickleback // Genetics. 2018.
- 55. McKinnon J.S., Rundle H.D. Speciation in nature: the threespine stickleback model systems // Trends in Ecology & Evolution. 2002. Vol. 17, № 10. P. 480–488.
- Hagen D.W. Isolating mechanisms in threespine sticklebacks (Gasterosteus) // J. Fish. Res. Bd. Can. 1967. Vol. 24, № 8. P. 1637–1692.
- Bell M.A., Foster S.A. The evolutionary biology of the threespine stickleback. 1st ed. Oxford: Oxford Science Publications, 1994.
- 58. Bell M.A. Intraspecific systematics of Gasterosteus aculeatus populations: implications for behavioral ecology // Behaviour. 1995. Vol. 132, № 15/16. P. 1131–1152.
- 59. Colosimo P.F. et al. Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles // Science. 2005. Vol. 307, № 5717. P. 1928.
- 60. Nelson T., Cresko W. Ancient genomic variation underlies repeated ecological adaptation in young stickleback populations // Evolution Letters. 2018. Vol. 2, № 1. P. 9–21.
- 61. Schluter D., Conte G.L. Genetics and ecological speciation // Proceedings of the National Academy of Sciences. 2009. Vol. 106, № Supplement 1. P. 9955–9962.

- 62. Leffler E.M. et al. Multiple Instances of Ancient Balancing Selection Shared Between Humans and Chimpanzees // Science. 2013. Vol. 339, № 6127. P. 1578.
- 63. Hohenlohe P.A. et al. Population Genomics of Parallel Adaptation in Threespine Stickleback using Sequenced RAD Tags // PLOS Genetics. 2010. Vol. 6, № 2. P. e1000862.
- Bell M.A. Lateral plate evolution in the threespine stickleback: getting nowhere fast. // Genetica.
   2001. Vol. 112. P. 445–461.
- 65. Loehr J. et al. Heritability of asymmetry and lateral plate number in the threespine stickleback // PLOS ONE. 2012. Vol. 7, № 7. P. e39843.
- 66. O'Brown N.M. et al. A recurrent regulatory change underlying altered expression and Wnt response of the stickleback armor plates gene EDA // eLife / ed. Krumlauf R. 2015. Vol. 4. P. e05290.
- 67. Jones F.C. et al. Reproductive isolation in a threespine stickleback hybrid zone // Journal of Evolutionary Biology. 2006. Vol. 19, № 5. P. 1531–1544.
- 68. Furin C.G., von Hippel F.A., Bell M.A. Partial reproductive isolation of a recently derived residentfreshwater population of threespine stickleback (Gasterosteus Aculeatus) from its putative anadromous ancestor // Evolution. 2012. Vol. 66, № 10. P. 3277–3286.
- Ziuganov V.V. et al. Genetically isolated sympatric forms of threespine stickleback, Gasterosteus aculeatus, in Lake Azabachije (Kamchatka-peninsula, USSR). // Environ Biol Fish. Vol. 18. P. 241–247.
- Karve A.D., von Hippel F.A., Bell M.A. Isolation between sympatric anadromous and resident threespine stickleback species in Mud Lake, Alaska. // Environ Biol Fishes. 2008. Vol. 81. P. 287– 296.
- Ziuganov V.V. Factors determining morphological differentiation in Gasterosteus aculeatus (Pisces, Gasterosteidae) (in Russian). // Zool Zhurn. Vol. 57. P. 1686–1694.
- Ziuganov V.V. Genetics of osteal plate polymorphism and microevolution of threespine stickleback (Gasterosteus aculeatus L.) // Theoretical and Applied Genetics. Vol. 65. P. 239–246.
- 73. Clark P.U. et al. The Last Glacial Maximum // Science. 2009. Vol. 325, № 5941. P. 710.
- 74. Corner G.D. et al. Postglacial relative sea-level change and stratigraphy of raised coastal basins on Kola Peninsula, northwest Russia // Global and Planetary Change. 2001. Vol. 31, № 1. P. 155–177.
- 75. Kolka V.V., Korsakova O.P. Application of geological methods for dating of stone labyrinths on the White Sea coast. // Proceedings of the MSTU. 2005. Vol. 15. P. 349–356.
- 76. Sodeland M. et al. "Islands of divergence" in the atlantic cod genome represent polymorphic chromosomal rearrangements // Genome Biology and Evolution. 2016. Vol. 8, № 4. P. 1012–1022.

- 77. Malinsky M. et al. Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake // Science. 2015. Vol. 350, № 6267. P. 1493–1498.
- 78. Ma T. et al. Ancient polymorphisms and divergence hitchhiking contribute to genomic islands of divergence within a poplar species complex // Proceedings of the National Academy of Sciences of the United States of America. 2018. Vol. 115, № 2. P. E236–E243.
- 79. Blanchette M. et al. Aligning multiple genomic sequences with the threaded blockset aligner // Genome Research. 2004. Vol. 14, № 4. P. 708–715.
- 80. Capra J.A. et al. A model-based analysis of GC-biased gene conversion in the human and chimpanzee genomes // PLOS Genetics. 2013. Vol. 9, № 8. P. e1003684.
- Lindblad-Toh K. et al. A high-resolution map of human evolutionary constraint using 29 mammals // Nature. 2011. Vol. 478. P. 476.
- Dixon J.R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions // Nature. 2012. Vol. 485. P. 376.
- 83. The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes // Nature. 2012. Vol. 491. P. 56.
- 84. Wong M.K.-S. et al. Flexible selection of diversified Na+/K+-ATPase α-subunit isoforms for osmoregulation in teleosts // Zoological Letters. 2016. Vol. 2, № 1. P. 15.
- 85. Auton A. et al. A fine-scale chimpanzee genetic map from population sequencing // Science. 2012.
   Vol. 336, № 6078. P. 193.
- 86. Brunschwig H. et al. Fine-scale maps of recombination rates and hotspots in the mouse genome // Genetics. 2012. Vol. 191, № 3. P. 757.
- 87. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood // Molecular Biology and Evolution.
  2007. Vol. 24, № 8. P. 1586–1591.
- 88. Lynch M. Evolution of the mutation rate // Trends in Genetics. 2010. Vol. 26, № 8. P. 345–352.
- Carbone L. et al. Gibbon genome and the fast karyotype evolution of small apes // Nature. 2014. Vol. 513. P. 195.
- 90. Glémin S. et al. Quantification of GC-biased gene conversion in the human genome // Genome Research. 2015. Vol. 25, № 8. P. 1215–1228.
- 91. Duret L., Galtier N. Biased gene conversion and the evolution of mammalian genomic landscapes // Annu. Rev. Genom. Hum. Genet. 2009. Vol. 10, № 1. P. 285–311.
- 92. dos Reis M. et al. Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny // Proc Biol Sci. 2012. Vol. 279, № 1742. P. 3491.

- 93. Cain C.E. et al. Gene expression differences among primates are associated with changes in a histone epigenetic modification // Genetics. 2011. Vol. 187, № 4. P. 1225.
- 94. Zhou X. et al. Epigenetic modifications are associated with inter-species gene expression variation in primates // Genome Biology. 2014. Vol. 15, № 12. P. 547.
- 95. Lercher M.J., Hurst L.D. Human SNP variability and mutation rate are higher in regions of high recombination // Trends in Genetics. 2002. Vol. 18, № 7. P. 337–340.
- 96. Webster M.T., Hurst L.D. Direct and indirect consequences of meiotic recombination: implications for genome evolution // Trends in Genetics. 2012. Vol. 28, № 3. P. 101–109.
- Yang S. et al. Parent–progeny sequencing indicates higher mutation rates in heterozygotes // Nature.
   2015. Vol. 523. P. 463.
- 98. Roesti M., Moser D., Berner D. Recombination in the threespine stickleback genome—patterns and consequences // Molecular Ecology. 2013. Vol. 22, № 11. P. 3014–3027.
- 99. DeFaveri J., Merilä J. Variation in Age and Size in Fennoscandian Three-Spined Sticklebacks (Gasterosteus aculeatus) // PLOS ONE. 2013. Vol. 8, № 11. P. e80866.
- 100. Gillespie J.H. Population genetics: a concise guide. Johns Hopkins University Press, 2004. 214 p.
- 101. Natri H.M., Shikano T., Merilä J. Progressive Recombination Suppression and Differentiation in Recently Evolved Neo-sex Chromosomes // Molecular Biology and Evolution. 2013. Vol. 30, № 5. P. 1131–1144.
- 102. Samuk K. et al. Gene flow and selection interact to promote adaptive divergence in regions of low recombination // Molecular Ecology. 2017. Vol. 26, № 17. P. 4378–4390.
- 103. Smith J.M., Haigh J. The hitch-hiking effect of a favourable gene. // Genet Res. 1974. Vol. 23, №
  1. P. 23–35.
- 104. Hohenlohe P.A. et al. Extensive linkage disequilibrium and parallel adaptive divergence across threespine stickleback genomes // Philos Trans R Soc Lond B Biol Sci. 2011. Vol. 367, № 1587. P. 395.
- 105. Hermisson J., Pennings P.S. Soft sweeps // Genetics. 2005. Vol. 169, № 4. P. 2335.
- 106. Messer P.W., Petrov D.A. Population genomics of rapid adaptation by soft selective sweeps // Trends in ecology & evolution. 2013. Vol. 28, № 11.
- 107. Roesti M. et al. The genomic signature of parallel adaptation from shared genetic variation // Molecular Ecology. 2014. Vol. 23, № 16. P. 3944–3956.
- McDonald J.H., Kreitman M. Adaptive protein evolution at the Adh locus in Drosophila // Nature.
   1991. Vol. 351. P. 652.

- 109. Dalziel A.C. et al. Origins and functional diversification of salinity-responsive Na(+), K(+) ATPase α1 paralogs in salmonids // Mol Ecol. 2014. Vol. 23.
- 110. Feng S.H. et al. Gene expression of Na+-K+-ATPase alpha 1 and alpha 3 subunits in gills of the teleost Oreochromis mossambicus, adapted to different environmental salinities // Mar Biotechnol (NY). 2002. Vol. 4.
- 111. Neerincx A. et al. A Role for the Human Nucleotide-binding Domain, Leucine-rich Repeatcontaining Family Member NLRC5 in Antiviral Responses // Journal of Biological Chemistry. 2010.
   Vol. 285, № 34. P. 26223–26232.
- 112. Welt C. et al. Activins, Inhibins, and Follistatins: From Endocrinology to Signaling. A Paradigm for the New Millennium // Exp Biol Med (Maywood). 2002. Vol. 227, № 9. P. 724–752.
- 113. Aruga J., Yokota N., Mikoshiba K. Human SLITRK family genes: genomic organization and expression profiling in normal brain and brain tumor tissue // Gene. 2003. Vol. 315. P. 87–94.
- Park C. et al. Deletion in Catna2, encoding αN-catenin, causes cerebellar and hippocampal lamination defects and impaired startle modulation // Nature Genetics. 2002. Vol. 31. P. 279.
- 115. Wagner G.F., Jaworski E.M., Haddad M. Stanniocalcin in the seawater salmon: structure, function, and regulation // American Journal of Physiology-Regulatory, Integrative and Comparative Physiology. 1998. Vol. 274, № 4. P. R1177–R1185.
- 116. Thompson A.J., Lummis S.C.R. 5-HT3 Receptors // Current Pharmaceutical Design. 2006. Vol.
  12, № 28. P. 3615–3630.
- Seear P.J. et al. The molecular evolution of spiggin nesting glue in sticklebacks // Molecular
   Ecology. 2015. Vol. 24, № 17. P. 4474–4488.
- 118. Klion A.D., Nutman T.B. The role of eosinophils in host defense against helminth parasites // Journal of Allergy and Clinical Immunology. 2004. Vol. 113, № 1. P. 30–37.
- Chiu S.-L., Cline H.T. Insulin receptor signaling in the development of neuronal structure and function // Neural Development. 2010. Vol. 5, № 1. P. 7.
- Rother K.I., Accili D. Role of insulin receptors and IGF receptors in growth and development // Pediatric Nephrology. Vol. 14. P. 558–561.
- 121. Haldane J.B.S. A mathematical theory of natural and artificial selection. Part 1. // Transactions of the Cambridge philosophical society Trinity College. P. 19–41.
- 122. Miller C.T. et al. Modular Skeletal Evolution in Sticklebacks Is Controlled by Additive and Clustered Quantitative Trait Loci // Genetics. 2014. Vol. 197, № 1. P. 405.

- 123. McCairns R.J.S., Bernatchez L. Adaptive divergence between freshwater and marine sticklebacks: insights into the role of phenotypic plasticity from an integrated analysis of candidate gene expression // Evolution. 2010. Vol. 64, № 4. P. 1029–1047.
- 124. Buhi W. Characterization and biological roles of oviduct-specific, oestrogen-dependent glycoprotein // Reproduction. 2002. Vol. 123, № 3. P. 355–362.
- 125. Cook L.M. The Rise and Fall of the Carbonaria Form of the Peppered Moth // The Quarterly Review of Biology. 2003. Vol. 78, № 4. P. 399–417.
- 126. Bersaglieri T. et al. Genetic Signatures of Strong Recent Positive Selection at the Lactase Gene // The American Journal of Human Genetics. 2004. Vol. 74, № 6. P. 1111–1120.
- 127. Rockman M.V. et al. Ancient and Recent Positive Selection Transformed Opioid cis-Regulation in Humans // PLOS Biology. 2005. Vol. 3, № 12. P. e387.
- 128. Wilde S. et al. Direct evidence for positive selection of skin, hair, and eye pigmentation in Europeans during the last 5,000 y // Proc Natl Acad Sci USA. 2014. Vol. 111, № 13. P. 4832.
- 129. Faustova M. et al. Radiation of European Eubosmina (Cladocera) from Bosmina(E.) longispina—
   concordance of multipopulation molecular data with paleolimnology // Limnology and Oceanography.
   2011. Vol. 56, № 2. P. 440–450.
- Dimmick W.W., Berendzen P.B., Golubtsov A.S. Genetic Comparison of Three Barbus (Cyprinidae) Morphotypes from the Genale River, Ethiopia // Copeia. 2001. Vol. 2001, № 4. P. 1123– 1129.
- 131. Senchukova A.L. et al. Genetic differentiation of chars (Genus Salvelinus) from lake Kronotskoe based on analysis of mitochondrial DNA // Journal of Ichthyology. 2012. Vol. 52, № 6. P. 389–399.