

УЧРЕЖДЕНИЕ РОССИЙСКОЙ АКАДЕМИИ НАУК  
ИНСТИТУТ ПРОБЛЕМ ПЕРЕДАЧИ ИНФОРМАЦИИ ИМ. А.А.ХАРКЕВИЧА

На правах рукописи

Науменко Сергей Анатольевич

ДИНАМИКА ОДНОЛОКУСНОГО МУЛЬТИАЛЛЕЛЬНОГО  
АДАПТИВНОГО ЛАНДШАФТА В МОЛЕКУЛЯРНОЙ ЭВОЛЮЦИИ БЕЛОК-  
КОДИРУЮЩИХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ ДНК

03.01.09 — математическая биология, биоинформатика

Диссертация на соискание учёной степени  
кандидата биологических наук

Научный руководитель:  
кандидат биологических наук  
Г.А. Базыкин

Москва — 2014

## Оглавление

Введение.....	4
Объект исследования.....	4
Предмет исследования.....	7
Методы исследования.....	9
Актуальность темы.....	9
Цели и задачи исследования.....	10
Научная новизна и практическая значимость.....	11
Апробация работы.....	12
Структура и объем диссертации.....	12
Список публикаций по теме диссертации.....	13
1 Развитие концепции адаптивного ландшафта.....	16
1.1 Пространство комбинаций аллелей.....	18
1.2 Пространство частот генотипов в популяции.....	21
1.3 Адаптивный ландшафт с глобальным пиком.....	22
1.4 Адаптивный ландшафт над пространством последовательностей.....	24
1.5 Двойные замены на мутационном ландшафте.....	26
1.6 Адаптивный ландшафт М. Кимуры.....	28
1.7 Адаптивные прогулки на изломанных ландшафтах.....	29
1.8 НК-модель С. Кауффмана.....	31
1.9 Флуктуации отбора.....	32
1.10 Неравновесная динамика адаптивных ландшафтов.....	33
1.11 Дырявый ландшафт.....	35
1.12 Практическое применение теории адаптивных ландшафтов.....	36
1.13 Изучение ландшафтов по молекулярным данным.....	37
1.14 Адаптивные ландшафты в эволюции вирусов.....	38
1.15 Выводы.....	39
2 Динамика триаллельного адаптивного ландшафта.....	41
2.1 Прямые, обратные и боковые замены в триаллельном сайте.....	41
2.2 Случай увеличения приспособленности производного аллеля.....	43
2.3 Случай уменьшения приспособленности предкового аллеля.....	46
2.4 Выводы.....	48
3 Анализ динамики адаптивного ландшафта в эволюции позвоночных и насекомых.....	49
3.1 Филогенетический анализ динамики адаптивных ландшафтов.....	49
3.2 Динамика частот обратных и боковых замен.....	56
3.3 Выводы.....	64
4 Оценка эффекта гетерогенности адаптивного ландшафта между сайтами.....	65
4.1 Эффект гетерогенности сайтов.....	65
4.2 Две характеристики белок-кодирующего сайта.....	68
4.2.1 Скорость эволюции.....	70
4.2.2 Ширина спектра допустимых аминокислот.....	70
4.3 Методика расчета.....	71

4.3.1 Филогении и наборы данных.....	71
4.3.2 Реконструкция замен в предковых линиях.....	72
4.4 Полученные корреляции.....	74
4.4.1 Корреляции по сайтам.....	74
4.4.2 Корреляции с усреднением по белку.....	75
4.4.3 Корреляции по классам аминокислот.....	76
4.5 Выводы.....	77
5 Размещение стоп-кодона уменьшает количество нонсенс-мутаций.....	80
5.1 Возникновение генетического кода .....	80
5.2 Разметка генетического кода.....	82
5.3 Критерии адаптации для расположения стоп-кодонов.....	83
5.4 Положение разметки универсального генетического кода на ландшафте всех кодов.....	84
5.4.1 Устойчивость к мутациям сдвига рамки.....	84
5.4.2 Устойчивость к нонсенс-мутациям.....	85
5.5 Выводы.....	87
Основные результаты и выводы.....	88
Благодарности.....	90
Список литературы.....	91

## Введение

### Объект исследования

Последовательность белка определяется последовательностью белок-кодирующего гена в соответствии с генетическим кодом. Несинонимические замены в белок-кодирующем гене в ходе эволюции приводят к изменению последовательности белка, которая определяет то, как он выполняет свою функцию в клетке, что, в свою очередь, влияет на приспособленность организма. Таким образом, определенный аллель в определенном сайте белок-кодирующего гена дает некоторый вклад в приспособленность организма. Назовем этот вклад приспособленностью аллеля.

Теоретически в аминокислотном сайте допустимы 20 аллелей, соответствующие 20 аминокислотам. В природе в каждом сайте некоторые аминокислоты, как правило, недопустимы, так что каждый сайт обычно ограничен их подмножеством. Кроме того, обычно выделяют классы аминокислот, сходных по физико-химическим свойствам; если разные аминокислоты в пределах класса являются функционально эквивалентными, то в качестве аллеля может выступать класс аминокислот.

Аллели различаются по абсолютной приспособленности  $w_a$ ; эта величина может принимать значение от 0 (летальный аллель, наличие которого приводит к гибели организма) до некоторой положительной величины, характеризующей работу белка (скажем, скорость реакции, которую он катализирует).

На практике удобно определить абсолютную приспособленность аллеля в контексте популяции: приспособленность аллеля — это средняя приспособленность популяции, в которой этот аллель зафиксирован, а частоты аллелей во всех остальных локусах сохранены. Очевидно, что приспособленность одного и того же аллеля в разных популяциях может различаться.

Относительная приспособленность аллеля определяется по отношению к аллелю, обладающему максимальной приспособленностью:

$$w = \frac{w_a}{w_{max}}$$

Эта величина изменяется в промежутке  $[0,1]$ .

Относительная приспособленность аллеля зависит от генетического состава популяции: один и тот же аллель может иметь разную приспособленность на разном генетическом фоне.

В каждый момент времени аллели  $x_1, x_2, \dots, x_n$ , допустимые в данном сайте, обладают приспособленностями  $w(x_1), w(x_2), \dots, w(x_n)$ . В популяции могут расщепляться несколько аллелей (внутрипопуляционный полиморфизм), однако при рассмотрении молекулярной эволюции на уровне видов мы считаем, что в каждый момент времени данный вид характеризуется только одним аллелем в данном сайте.

Приспособленность аллеля, как присутствующего в сайте, так и отсутствующего, может меняться со временем,  $w = w(t)$ .

Адаптивный ландшафт (fitness landscape, или adaptive landscape) — это функция, сопоставляющая генотипу значение приспособленности. Генотип определяется набором аллелей в соответствующих локусах.

Адаптивный ландшафт для двух локусов  $L$  и  $M$  с некоторым набором аллелей  $\{l_1, \dots, l_n\}$ ,  $\{m_1, \dots, m_k\}$  в каждом из них можно изобразить в виде трехмерного графика (Рисунок 1), на котором для каждого из сочетаний аллелей  $l_i$  и  $m_j$  определена величина  $z(l_i, m_j)$  - приспособленность, которую дает данное сочетание признаков. Области высокой приспособленности называются адаптивными пиками, области низкой приспособленности — адаптивными долинами.

Если рассматривать эволюцию на масштабе генов или на масштабе полных геномов, размерность адаптивного ландшафта возрастает; он также может становиться изломанным — с множеством локальных пиков.

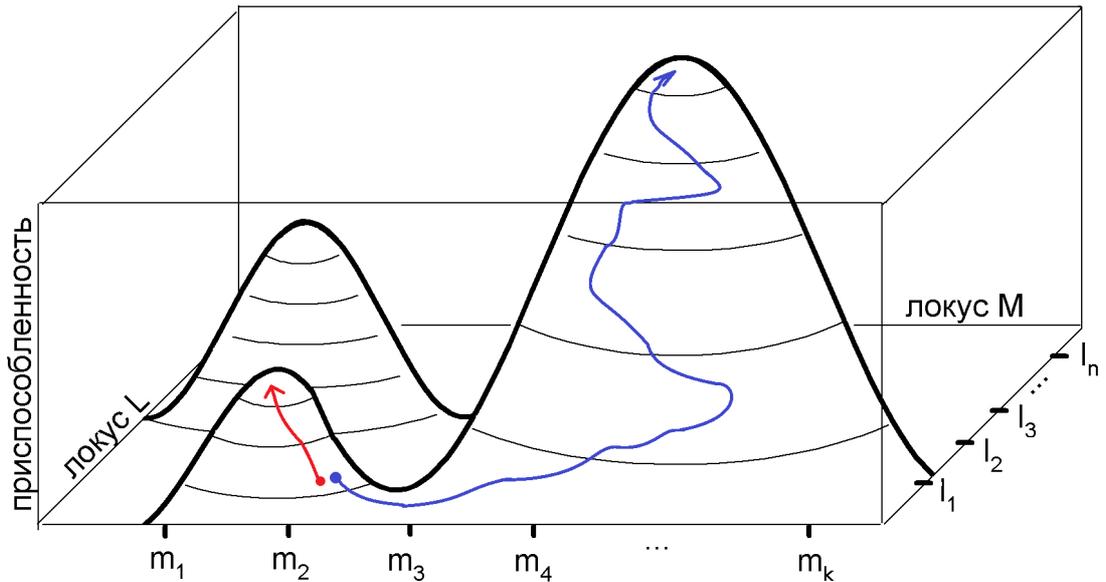


Рисунок 1. Адаптивный ландшафт в пространстве двух локусов. Изображение ландшафта в виде непрерывной поверхности условно; ландшафт является дискретным

Генотипы особей одного вида в популяции различаются; так, для человека среднее попарное нуклеотидное различие между двумя генотипами составляет 0,1%. Таким образом, на адаптивном ландшафте популяция особей будет отображаться как облако близко расположенных точек.

Эволюционный процесс можно представить себе как движение популяций по адаптивному ландшафту под действием основных факторов эволюции — отбора, мутаций, генетического дрейфа (Рисунок 1).

В данной работе исследуется адаптивный ландшафт аминокислотного сайта белок-кодирующего гена; соответственно, адаптивным ландшафтом называется вектор-функция  $\vec{f}$ , для каждого момента времени  $t$  задающая вектор значений приспособленности  $\vec{W} = (w(x_1), w(x_2), \dots, w(x_n))$  для всех аллелей  $\{x_1, \dots, x_n\}$  в данном сайте:

$$\vec{W} = \vec{f}(t).$$

Генетический код формально можно представить как отображение

имеющего определенную внутреннюю структуру множества 64 нуклеотидных триплетов (кодонов) на множество аминокислот. Триплеты делятся на смысловые, кодирующие аминокислоты, и стоп-кодоны (TAA, TAG, TGA). Назовем разметкой генетического кода способ деления кодонов на смысловые и стоп-кодоны. Генетические коды с различной разметкой могут обладать различной приспособленностью. В данной работе исследуется адаптивный ландшафт на пространстве разметок генетических кодов.

## Предмет исследования

Хотя нуклеотидную или аминокислотную последовательности можно рассматривать как простую линейную структуру, в клетке белок свернут в сложную структуру, в которой удаленные по линейному порядку аминокислоты оказываются рядом, и соотношение их свойств определяет выполнение белком своей функции. Кроме того, аминокислоты в различных положениях могут взаимодействовать посредством влияния на стабильность белка. Таким образом, замена одной аминокислоты влияет на приспособленность других аминокислот. Такие взаимодействия называются эпистатическими (epistatic interactions), в отличие от случая, когда аминокислоты в сайтах влияют на приспособленность белка независимо.

Адаптивный ландшафт аминокислотного сайта не является статическим: даже если в сайте не происходят замены, приспособленность аллелей в нем может меняться из-за замен в других сайтах за счет эпистатических взаимодействий.

В данной работе изучается динамика адаптивного ландшафта после замены одной аминокислоты (A) на другую (B). Рассматриваются отличия между видами, т.е. прямая замена происходит на макроэволюционных временах. Будем обозначать такую замену  $A \rightarrow B$ . Изучается вопрос о том, что происходит

с приспособленностью аллеля А после такой замены с течением времени (Рисунок 2).

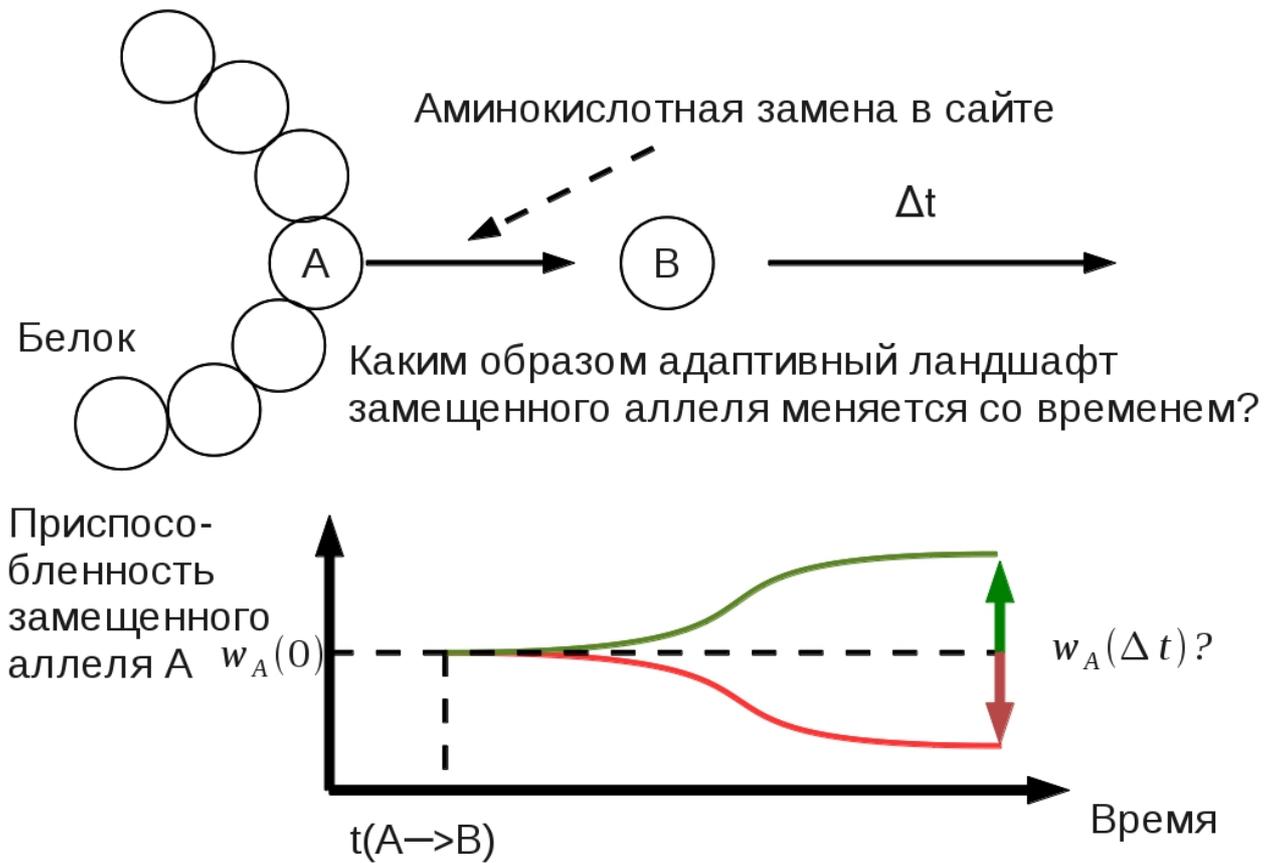


Рисунок 2. Локальный адаптивный ландшафт

Расположение стоп-кодонов в генетическом коде влияет на частоту точечных нонсенс-мутаций (мутаций, переводящих смысловой кодон в стоп-кодон) и нонсенс-мутаций при сдвиге рамки считывания. Таким образом, альтернативные генетические коды могут различаться по устойчивости к нонсенс-мутациям. В данной работе изучается устойчивость разметки стандартного генетического кода по сравнению с альтернативными разметками.

## **Методы исследования**

Непосредственное изучение адаптивного ландшафта возможно для отдельно взятых белков, когда путем замены отдельных аминокислот о приспособленности судят, например, по результатам моделирования пространственной структуры, или для генно-модифицированных модельных организмов (о приспособленности судят по скорости размножения организмов, несущих различные аллели).

Также изучение адаптивных ландшафтов возможно по данным популяционной генетики: аллели, которые наблюдаются чаще в популяции (имеют большую частоту), обладают большей приспособленностью, а наблюдаемые реже — меньшей.

В данной работе изучение динамики адаптивного ландшафта на больших эволюционных временах, для клады позвоночных и насекомых, проводится с использованием данных о внутрипопуляционной изменчивости, а также при помощи сравнительной геномики: рассчитывая частоты замен в молекулярной эволюции можно делать выводы о динамике ландшафта.

## **Актуальность темы**

Адаптивный ландшафт является одной из центральных концепций современной эволюционной биологии. Процесс эволюции рассматривается как движение популяций по адаптивному ландшафту в результате изменения генотипов.

Математический аппарат эволюционной биологии — популяционная генетика — складывался как изучение равновесных состояний на фоне факторов эволюции (мутаций, отбора, дрейфа), действующих с постоянной

силой. Однако, как становится ясно, все эти факторы меняются во времени, что вызывает необходимость обобщения классических представлений на подвижные адаптивные ландшафты.

В данной работе затрагивается тема динамики адаптивного ландшафта для одного из простейших случаев. Выбор масштаба объекта исследования позволяет полностью проанализировать динамику адаптивного ландшафта аналитически. С другой стороны, накопление молекулярных данных в масштабах полных геномов позволяет провести анализ динамики адаптивного ландшафта при помощи методов биоинформатики. Таким образом, удастся сравнить предсказания теоретических моделей с эволюцией молекулярных последовательностей.

Эволюция ранней жизни на Земле, в частности, вопрос о возникновении аппарата трансляции, остается сравнительно малоизученной темой. С открытием альтернативных генетических кодов актуализировался вопрос о возникновении и эволюции генетического кода. В данной работе изучается одна из первичных адаптаций аппарата трансляции — выделение стоп-кодона из множества смысловых.

## Цели и задачи исследования

Целью данной работы является изучение динамики однолокусного адаптивного ландшафта в эволюции белок-кодирующих генов позвоночных и насекомых. В ходе работы были решены следующие задачи:

- подготовлены множественные выравнивания белок-кодирующих генов для филогений позвоночных и насекомых, включающие информацию о полиморфизме у *Drosophila melanogaster* и *Homo sapiens*;
- восстановлены предковые состояния во внутренних узлах филогений;
- разработано программное обеспечение, позволяющее отбирать сайты из

- полногеномных выравниваний, соответствующие заданным шаблонам;
- рассчитаны частоты обратных замен в эволюции позвоночных и насекомых для различных эволюционных расстояний;
  - проведен теоретический анализ динамики однолокусного адаптивного ландшафта с тремя аллелями;
  - разработан ABC-тест, позволяющий оценить относительную силу эффекта роста приспособленности производного аллеля и эффекта уменьшения приспособленности предкового аллеля;
  - проведена оценка эффекта гетерогенности сайтов.

В ходе изучения первичных адаптаций аппарата трансляции было определено положение стандартного генетического кода на ландшафте всех возможных кодов для двух критериев приспособленности:

- вероятность прерывания трансляции мРНК со сдвигом рамки;
- вероятность точечной нонсенс-мутации.

### **Научная новизна и практическая значимость**

В этой работе был проведен теоретический и сравнительно-геномный филогенетический анализ динамики ландшафта приспособленности сайта белок-кодирующего гена.

Было показано, что в молекулярной эволюции происходит эффект «забывания», который состоит в том, что вклад в приспособленность ранее присутствовавших в сайте аллелей падает со временем до фонового значения. По-видимому, это подтверждает гипотезу о наличии множества эпистатических связей между сайтами белок-кодирующего гена.

Результаты работы важны для понимания общих закономерностей эволюционного процесса и могут быть использованы при изучении адаптивных

ландшафтов в молекулярной эволюции, в том числе для экспериментального изучения адаптивных ландшафтов в эволюции вирусов и бактерий.

### **Апробация работы**

По результатам работы были сделаны доклады на следующих конференциях: 13-й конференции “Математика. Компьютер. Образование.” (г. Дубна, 23-28 января 2006 г.), 32-й и 33-й конференциях “Информационные технологии и системы” (ИТиС'09: пос. Бекасово, 15-18 декабря 2009 г., ИТиС'10: г. Геленджик, 20-24 сентября 2010 г.), ежегодных конференциях сообщества по молекулярной биологии и эволюции (SMBE'10: г. Лион, 4-8 июля 2010 г., SMBE'12: г.Дублин, 23-26 июня), 2-й и 3-й конференциях “Математическая биология и биоинформатика” (г. Пущино, 7-13 сентября 2008 г., 10-15 октября 2010 г.), 3-й и 5-й конференциях молодых ученых “Биология: от молекулы до биосферы” (г. Харьков, 18-21 ноября 2008 г., 22-25 ноября 2010 г.), Московской конференции по вычислительной молекулярной биологии (МССМВ'11, г. Москва, 21-24 июля 2011 г.), конференции по образованию в биоинформатике (RECOMB'12 Satellite Conference on Bioinformatics Education, г. Санкт-Петербург, 26 августа 2012 г.).

### **Структура и объем диссертации**

Работа состоит из введения, пяти глав, заключения и списка литературы. Глава 1 содержит обзор литературы по теме диссертации. В главе 2 дается теоретический анализ динамики адаптивного ландшафта. В главах 3 и 4 изложены результаты анализа динамики адаптивного ландшафта и анализа взаимосвязи ширины спектра допустимых аминокислот и скорости молекулярной эволюции для белок-кодирующих генов позвоночных и

насекомых. В главе 5 содержится анализ адаптивного ландшафта разметок генетических кодов.

Работа изложена на 97 страницах, содержит 7 таблиц и 21 рисунок. Список литературы содержит 80 наименований.

## **Список публикаций по теме диссертации**

### **Статьи в научных журналах**

1. Naumenko S.A., Kondrashov A.S., Bazykin G.A. Fitness conferred by replaced amino acids declines with time. // *Biology Letters*. - 2012. - V. 8. - N 5 - P.825-828.
2. Naumenko S.A., Kondrashov A.S. Rate and breadth of protein evolution are only weakly correlated. // *Biology Direct*. - 2012. - V. 7. - N 8. - P.1-12.
3. Малинецкий Г.Г., Науменко С.А., Подлазов А.В. Об экстремальных свойствах разметки генетического кода. // Доклады академии наук (биохимия, биофизика, молекулярная биология). - 2007. - Т. 414. - N 6. - С.831-835.

### **Тезисы конференций**

1. Науменко С.А., Подлазов А.В. Об экстремальных свойствах разметки генетического кода. // “Математика. Компьютер. Образование.” Сборник трудов XIII международной конференции. Под общей редакцией Г.Ю. Ризниченко. - Ижевск: Научно-издательский центр “Регулярная и хаотическая динамика”. - 2006. - Т.2. - С.404.-413.
2. Науменко С.А. Роль нонсенс-кодонов в обеспечении оптимальности генетического кода. // Материалы докладов XV Международной конференции студентов, аспирантов и молодых ученых “Ломоносов” / Отв. ред. И.А. Алешковский, П.Н. Костылев. [Электронный ресурс] — М.: Издательство МГУ; СП МЫСЛЬ, 2008. — 1 электрон. опт. диск (CD-ROM); 12 см. - Систем. требования: ПК с процессором 486 +; Windows

- 95; дисковод CD-ROM; Adobe Acrobat Reader.
3. Науменко С.А., Кондрашов А.С., Базыкин Г.А. О высоких частотах реверсий в эволюции позвоночных и насекомых. // “Биология: от молекулы до биосферы”. Материалы IV Международной конференции молодых ученых (17-21 ноября 2009 г., г. Харьков, Украина). - Харьков: ЧПИ “Новое слово”. - 2009. - С.149- 150.
  4. Naumenko S., Kondrashov A., Bazykin G.A. High Frequency of reversals in evolution of vertebrates and insects // Информационные технологии и системы (ИТиС'09): сборник трудов конференции. [Электронный ресурс] - М.: ИППИ РАН. - 2009. - 463с. - 1 электрон. опт. диск (CD-ROM) - С.367- 368.
  5. Naumenko S., Kondrashov A., Bazykin G. High frequency of reversals in evolution of vertebrates and insects. // Annual meeting of the society for molecular biology and evolution “SMBE-2010”, Lyon, France, July 4-8.
  6. Naumenko S., Kondrashov A., Bazykin G. Frequency of reversals in evolution of vertebrates and insects decreases on increased phylogenetic distance between substitutions // Информационные технологии и системы (ИТиС'10): сборник трудов конференции. - [Электронный ресурс] — М.: ИППИ РАН. - 2010. - С.361- 362.
  7. Науменко С.А., Кондрашов А.С., Базыкин Г.А. Исследование эволюционных ландшафтов приспособленности аминокислотных сайтов при помощи реверсий // Математическая биология и биоинформатика: III Международная конференция., г. Пущино, 10-15 октября 2010 г.: Доклады / Под. ред. В.Д. Лахно. - С.105-106.
  8. Науменко С.А., Базыкин Г.А. Вклад в приспособленность недавно замененной аминокислоты уменьшается со временем. // “Биология: от молекулы до биосферы”. Материалы V Международной конференции молодых ученых (22-25 ноября 2010 г., г.Харьков, Украина). - Х.: Оперативная полиграфия. - 2010. - С.9-10.

9. Naumenko S.A., Kondrashov A.S., Bazykin G.A. The fitness conferred by recently replaced amino acids rapidly declines with time. // Proceedings of the International Moscow Conference on Computational Molecular Biology (MCCMB'11). - 2011. - P.243.
10. Naumenko S.A., Kondrashov A.S. Rate and breadth of protein evolution are only weakly correlated. // Annual meeting of the society for molecular biology and evolution "SMBE-2012", Dublin, Ireland, June 23-26.
11. Naumenko S.A. The number of reversing substitutions. // The program of Forth RECOMB Satellite Conference on Bioinformatics Education (RECOMB-BE 2012). - 2012. - P.11.

## 1 Развитие концепции адаптивного ландшафта

Адаптивный ландшафт — это функция, сопоставляющая генотипу значение приспособленности.

Концепция адаптивного ландшафта прочно вошла в учебники эволюционной биологии и популяционной генетики (Таблица 1.1).

Таблица 1.1. Использование концепции адаптивного ландшафта в учебниках

Где используется	Тема
Hamilton M.B. Population genetics. [1]	Отбор в различных системах доминирования, отбор в триаллельном локусе, отбор в двух диаллельных локусах, эволюционные сценарии С. Райта (S. Wright), критика теории смещенного равновесия Райта.
Crow J.F., Kimura M. An introduction to population genetics theory. [2]	Принцип максимума для направленного отбора.
Hartl D.L., Clark A.G. Principles of Population Genetics. [3]	Пересечение адаптивной долины за счет дрейфа, двухлокусная диаллельная модель, теория смещенного равновесия.
Bergstrom C.T., Dugatkin L.A. Evolution. [4]	Ландшафты в пространстве фенотипов, генотипов, компенсаторная замена.
Ridley M. Evolution. [5]	Ландшафт С. Райта, доминирование, двухлокусная модель, теория смещенного равновесия.
Futuyma D.J. Evolution. [6]	Отбор в различных системах доминирования, пересечение адаптивной долины за счет дрейфа, перипатрическое видообразование.
Kondrashov A.S. Evolution of Life. [7]	Связность генотипов и фенотипов, субоптимальность, гомология, отбор, мутации, эпистаз, генетический дрейф, гомоплазии, видообразование, адаптация. Простые и сложные признаки. Популяции на адаптивном ландшафте. Масштабы адаптивного ландшафта. Режимы отбора. Отбор, действующий на количественный признак. Переключение режимов отбора.

История развития представлений об адаптивном ландшафте изложена в обзорах, перечисленных в Таблице 1.2.

Таблица 1.2. Обзоры по адаптивным ландшафтам

Источник	Особенности изложения
Orr H.A. Genetic theories of adaptation: a brief history. [8]	Использование ландшафтов для моделирования процесса адаптации.
Gavrilets S. Fitness landscapes and the origin of species. [9]	Ландшафты в видообразовании.
Orr H.A. Fitness and its role in evolutionary genetics. [10]	Ландшафты и приспособленность.
Svensson E. and Calsbeek R. (eds). The adaptive landscape in evolutionary biology. [11].	В основном фенотипические ландшафты.

В биологической литературе насчитываются тысячи статей с использованием концепции адаптивного ландшафта. Адаптивные ландшафты рассматриваются в популяционной генетике, эволюционной экологии, количественной генетике, экспериментальной эволюции, теории сохранения биоразнообразия, теории видообразования и макроэволюционной динамики, в исследованиях по мимикрии, молекулярной биологии и биологии развития [11, с. 299].

Адаптивный ландшафт не является одной строго определенной моделью, а скорее представляет собой концепцию, методологию построения и верификации моделей эволюционных процессов – адаптации и видообразования.

Весь эволюционный процесс в целом – это любые движения популяций по адаптивному ландшафту под действием факторов эволюции – отбора, мутаций, генетического дрейфа, миграции, структуры и численности популяции.

Процесс адаптации представляется как движение по адаптивному ландшафту по градиенту приспособленности под действием отбора. Процесс видообразования можно представлять себе как разделение кластера точек

высокой приспособленности в пространстве генотипов, соответствующих популяции, на два независимых подкластера, которые удаляются друг от друга до тех пор, пока их генотипы не станут столь далекими, что это будет препятствовать скрещиванию.

Однолокусный адаптивный ландшафт можно представить в виде двухмерного графика функции приспособленности, а двухлокусный – в виде трехмерного графика. Однако в реальности даже простые биологические функции определяются значительно большим количеством локусов. Таким образом, адаптивный ландшафт является многомерным.

Невозможно построить точную модель адаптивного ландшафта на масштабе геномов. Даже в случае геномов небольшой длины — геномов бактерий, длина которых составляет порядка 3 млн. нуклеотидов, количество аллельных вариантов (генотипов) составит  $4^{3000000}$ , т.е. адаптивный ландшафт нужно рассматривать в 3 000 001-мерном пространстве. В эксперименте, наблюдая приспособленность популяций организмов, мы можем измерить приспособленность лишь для некоторых генотипов из этого большого пространства. Таким образом, можно наблюдать лишь некоторые из адаптивных пиков. Вся карта адаптивного ландшафта, включая и альтернативные адаптивные пики, остается скрытой от наблюдателя, равно как и от эволюционного процесса, который нащупывает пути к адаптивным пикам вслепую [7].

## 1.1 Пространство комбинаций аллелей

С. Райт заложил основы популяционной генетики, рассмотрев влияние мутаций, отбора и миграции на распределение частот генотипов в популяции [12]. Он использовал концепцию адаптивного ландшафта для иллюстрации своих результатов на шестом генетическом конгрессе [13]. С тех

пор эти две статьи в сумме были процитированы более восьми тысяч раз; за это время адаптивный ландшафт стал одной из важнейших концепций эволюционной биологии. В конце жизни С. Райт вернулся к теме адаптивного ландшафта [14], очертив область применения своих моделей эволюции по отношению к результатам Р. Фишера (R. Fisher), Дж. Холдейна (J. Haldane), М. Кимуры (M. Kimura) : С. Райт рассматривал случаи отклонения от панмиксии, в результате которых возникает конкуренция между обособленными группами, принадлежащими к одной популяции. К 80- летию статьи С. Райта было выпущено специальное издание, посвященное роли адаптивного ландшафта в различных разделах эволюционной биологии [11].

Ландшафт С. Райта имеет размерность  $N+1$ , где  $N$  - количество локусов в рассматриваемом генотипе. По каждой размерности перечислены аллели, допустимые в данном локусе. Дополнительная размерность задает приспособленность комбинаций аллелей. Некоторые сочетания аллелей дают высокую приспособленность (адаптивные пики), другие – низкую (адаптивные долины). Райт использовал условное трехмерное изображение таких пространств.

В саму модель С. Райта заложен эпистаз – неаддитивный режим влияния комбинаций аллелей на приспособленность. Это определение эпистаза С. Райт модифицировал из более ранних работ У. Бейтсона (W. Bateson), который понимал под эпистазом результат дигибридного скрещивания, дающий новый фенотип, в противоположность гипостазу – летальной комбинации аллелей. Отсутствие эпистаза по С. Райту не обязательно влечет за собой отсутствие эпистаза по У. Бейтсону [15].

С. Райт склонялся к мысли об изломанном ландшафте с большим количеством пиков и исследовал факторы, препятствующие процессу адаптации. Если процесс адаптации представить как подъем популяции вверх к адаптивному пику под действием естественного отбора, то при наличии

нескольких пиков популяция достигнет вершины того из них, который оказался ближе к её начальному положению. Этот пик не обязательно должен быть самым высоким, однако спуститься с него популяция уже не может, поскольку этому препятствует отрицательный отбор. Таким образом, популяция не достигает максимального уровня приспособленности, возможного на данном адаптивном ландшафте, и возникает проблема пересечения адаптивных долин, которая затрудняет построение реалистичных моделей адаптации и видообразования, ведь в природе мы наблюдаем высоко приспособленные виды.

Для преодоления этого противоречия С. Райт создал теорию смещенного равновесия: помимо отбора, в процессе эволюции действуют силы, которые могут привести к смещению равновесия. Райт выделял три стадии процесса эволюции. На начальном этапе глобальная популяция под действием генетического дрейфа делится на полу-изолированные подмножества — субпопуляции (демы). Такие субпопуляции будут иметь меньший уровень изменчивости, чем глобальная популяция, а также во многих из них средний уровень приспособленности будет ниже среднего уровня приспособленности в глобальной популяции. Однако субпопуляции будут находиться в разных точках адаптивного ландшафта, по соседству с разными локальными адаптивными пиками. На втором этапе под действием отбора приспособленность субпопуляций растет — они поднимаются на свои адаптивные пики. В субпопуляции за счет меньшей численности большую роль играют случайные процессы — генетический дрейф, в результате действия которого субпопуляция может пересечь адаптивную долину. На третьем этапе за счет отбора между субпопуляциями растет средняя приспособленность глобальной популяции: популяция с большей средней приспособленностью порождает больше мигрантов, которые, присоединяясь к равновесной популяции, могут сдвигать установившееся в ней равновесие за счет своей

высокой приспособленности и нетипичного для популяции набора аллелей.

В целом, теория смещенного равновесия С. Райта, как основное объяснение эволюционного процесса, не нашла подтверждения; однако это не сказалось на популярности концепции адаптивного ландшафта [10].

Для проверки теории смещенного равновесия было проведено множество опытов с искусственным отбором, в которых эволюция глобальной популяции сравнивалась с эволюцией популяции, подразделенной на группы. Большинство экспериментов показало более быстрый рост приспособленности для глобальной популяции по сравнению с подразделенной, т.е. противоречило предсказаниям теории смещенного равновесия; однако эти эксперименты предпринимались для признаков, приспособленность которых зависит аддитивно от комбинаций генов, а не эпистатически. С. Райт же строил свою теорию именно для эпистатического случая [11, с. 66].

Следует отметить, что С. Райт предполагал, что адаптивный ландшафт может меняться в результате изменения среды, а также уделял большое внимание эпистатическим взаимодействиям между генами. Изменение ландшафта также облегчает для популяций пересечение адаптивных долин.

## 1.2 Пространство частот генотипов в популяции

Р. Левонтин (R. Lewontin) и К. Койима (K. Kojima) на основе данных М. Уайта (M. White) о хромосомных инверсиях у австралийских короткоусых прямокрылых *Moraba scurra* впервые построили эмпирический адаптивный ландшафт на пространстве частот генотипов в популяциях для двухлокусного случая [16-18]. В этой модели эволюции адаптивный ландшафт представляет собой трехмерную конструкцию; каждая точка  $(X, Y, Z)$  соответствует средней приспособленности  $Z$  в популяции с заданными частотами  $(X, Y)$  аллелей в двух локусах. Анализ данного ландшафта позволил сделать важное наблюдение о

наличии эпистаза двух локусов с инверсионным полиморфизмом.

С. Райт упоминал о ландшафте частотного типа в своем сообщении Р. Фишеру, а также в работе, представленной во французский сборник по статистической биологии [11, с.9], однако в его основные работы эта модель не вошла.

С. Гаврилец (S. Gavrilets) отмечает, что ландшафт частотного типа (тип 2) может быть получен из ландшафта, построенного на пространстве генотипов (тип 1): в ландшафте 1-го типа группы точек соответствуют популяциям, следовательно, для каждой популяции могут быть рассчитаны частоты генотипов и средняя приспособленность [9, с. 32].

При увеличении количества локусов ландшафт 2-го типа приближается к ландшафту 1-го типа: количество возможных генотипов растет экспоненциально, при этом частота большинства из них равна нулю, следовательно, популяции характеризуются частотами присутствующих генотипов, что является особенностью ландшафта 1 типа.

### **1.3 Адаптивный ландшафт с глобальным пиком**

Р. Фишер предполагал, что при росте размерности адаптивного ландшафта (при увеличении количества рассматриваемых локусов) локальные адаптивные пики становятся седловыми точками, и существует единственная комбинация аллелей, дающая максимальную приспособленность, что позволяет популяции достичь глобального пика благодаря естественному отбору без действия других факторов эволюции [19; 9, с. 36; 5, с.218].

Фундаментальная теорема Р. Фишера и соответствующий ей адаптивный ландшафт известны в основном в интерпретации С. Райта, как случай локальной адаптации в отсутствие эпистаза, или как медленная адаптация в

большой панмиктической популяции. Модель адаптивного ландшафта Р. Фишера находит свое применение в некоторых специальных случаях: при аддитивной и мультипликативной схеме вклада аллелей в приспособленность (отсутствие эпистаза), в случае, когда приспособленность зависит от количества вредных аллелей, при изучении адаптации в окрестности локального пика [9, с.38].

Однако в действительности результат Р. Фишера гораздо сильнее [20,21,22]. Цель Р. Фишера состояла в том, чтобы изолировать отбор от всех остальных факторов эволюционного процесса и построить законы сохранения. Эти законы действуют для любой реальной популяции, однако, чтобы предсказывать долговременную динамику такой популяции (в чем заключалась цель С. Райта), необходимо учитывать все остальные факторы.

Первый из законов Р. Фишера — инвариантность скорости увеличения средней приспособленности в популяции за счет естественного отбора, которая пропорциональна генетическому разнообразию в популяции. Под инвариантностью понимается независимость этого соотношения от многих конкретных деталей генетики и экологии популяции: особенностей скрещивания, доминирования, факторов среды. Р. Фишер включал особенности доминирования и эпистаз в факторы среды. Под отбором Р. Фишер понимал выживание наиболее приспособленных при дифференциальном размножении, не включая в отборные эффекты изменение средней приспособленности в популяции за счет изменения среды (адаптацию).

Второй закон Р. Фишера заключается в том, что средняя скорость роста популяции (приспособленность) в любой момент времени близка к нулю. В случае отклонений этого значения от нуля либо численность популяции растет до бесконечности, либо популяция вымирает. Любое увеличение приспособленности за счет отбора компенсируется уменьшением за счет действия изменений среды (фактор обеднения среды): например, уменьшаются ресурсы или возрастает внутривидовая или межвидовая конкуренция.

Р. Фишер полагал, что флуктуирующий отбор приводит к появлению новых комбинаций аллелей на низких частотах с большей скоростью, чем генетический дрейф, а затем за счет отбора эти комбинации распространяются по популяции. Следовательно, для объяснения быстрой адаптации и разнообразия подразделение популяции на демы по С. Райту излишне.

#### **1.4 Адаптивный ландшафт над пространством последовательностей**

Дж.М. Смит (J. M. Smith) предложил модель эволюции как движение в дискретном пространстве белковых последовательностей [23]. Статья не содержит понятия “адаптивный ландшафт”, однако вводит пространство белковых последовательностей, между которыми возможны переходы, а также отмечает, что есть последовательности, имеющие смысл (биологический) и не имеющие такового, и что некоторые из последовательностей дают большую приспособленность в каких-то условиях. Фактически Дж.М. Смит описывает пространство последовательностей и некоторую функцию приспособленности. Кроме этого, говоря о возможной низкой приспособленности промежуточных вариантов, соседствующих с двумя высокоприспособленными, Дж.М. Смит описывает изломанность ландшафта, не давая самого термина. Белковые последовательности формируют сеть (связный ландшафт), если для любого белка найдется хотя бы один, достижимый одинарной заменой, приспособленность которого больше или равна приспособленности данного белка. Вероятность найти белок, имеющий биологический смысл, вблизи другого такого белка выше, чем среди случайных последовательностей. Для заданного белка длиной  $L=100$  аминокислот достижимыми за 1 мутацию являются примерно 1000 белков (из-за ограничений генетического кода). Если ландшафт является связным, то для заданного белка вероятность найти эволюционно доступный вариант должна быть больше  $1/1000$ , следовательно, пространство последовательностей имеет уплотнения – области белков с

высокой приспособленностью.

В модели Дж.М. Смита приспособленность белковой последовательности не меняется со временем.

И. Поволоцкая (I. Povolotskaya) и Ф. Кондрашов (F. Kondrashov) [24] показали, что расхождение белковых последовательностей в глобальном пространстве все еще происходит, несмотря на  $\sim 3,5 \cdot 10^9$  лет белковой эволюции. Действительно, количество аминокислотных замен в современных последовательностях, которые удаляют их от гомологичных древнейших последовательностей (вероятно, присутствовавших в первичных организмах (LUCA)), больше, чем количество замен, приближающих современные последовательности к предковым. Это верно для любых расстояний между современными и древнейшими последовательностями.

Столь медленная скорость белковой эволюции на адаптивном ландшафте может объясняться её эпистатическим характером: вклад в приспособленность аминокислоты в некотором сайте зависит от наличия определенных аминокислот в других сайтах, что делает адаптивный ландшафт сильно изломанным, а передвижение по нему — замедленным.

Падение относительной (нормированной на дивергенцию в синонимических сайтах) скорости конвергентной эволюции с расстоянием между современной и древней последовательностями, и практически неизменное значение скорости для дивергентной эволюции, подтверждают эпистатический характер процесса (24, Рисунок 5).

М. Брин (M. Breen) и соавторы [25] на наборах ортологичных последовательностей для 13 митохондриальных, 1 хлоропластного и 2 ядерных генов, каждый из которых состоял из более 1000 последовательностей, показали, что среднее количество допустимых аминокислот в сайте равно восьми. Следовательно, в отсутствие эпистатических ограничений можно ожидать средней скорости несинонимической эволюции, равной  $8/20 = 0,4$  от скорости нейтральной эволюции. В действительности скорость

несинонимической эволюции на порядок меньше, чем ожидается в отсутствие эпистаза. Это еще один аргумент в пользу представлений об изломанном адаптивном ландшафте.

### 1.5 Двойные замены на мутационном ландшафте

Дж. Гиллеспи (J. Gillespie) рассмотрел модель эволюции на адаптивном ландшафте в пространстве нуклеотидных последовательностей ДНК под действием отбора [26; 27, с.241].

Аллель в локусе длиной  $D$  нуклеотидов имеет  $3D$  соседей, достижимых за один мутационный шаг. Пусть в локусе нет нейтральных аллелей – все аллели обладают разной приспособленностью. В начальный момент времени в локусе зафиксирован аллель с максимальной приспособленностью из  $3D+1$  окрестности аллелей. В пространстве последовательностей существуют более приспособленные аллели, однако они находятся на расстояниях двух и более мутационных шагов. Частота одиночных нуклеотидных мутаций составляет около  $10^{-8}$ - $10^{-9}$  событий на нуклеотидный сайт в год, поэтому скорость движения популяции в пространстве генотипов мала, и вероятность фиксации аллеля на расстоянии двух и более замен от текущего близка к нулю.

В теории Дж. Гиллеспи аллели, которые в текущий момент расщепляются в популяции, уже обладают высокой приспособленностью за счет предыдущей адаптации, поэтому новые аллели отбираются из хвоста распределения приспособленностей (коэффициентов отбора), что позволяет использовать теорию экстремальных значений и получить результаты, независимые от конкретного вида распределения.

В результате изменения среды режим отбора меняется, и в  $3D$ -окрестности текущего аллеля может возникнуть более приспособленный аллель. Если он начнет фиксироваться в популяции, то для отбора станут доступными еще

3D - 1 новых аллелей, среди которых, в свою очередь, может оказаться аллель с еще большей приспособленностью.

Дж. Гиллеспи делает вывод, что движение по ландшафту происходит в виде “вспышек”, т.е. быстрых последовательностей из нескольких замен, и оценивает длину такой вспышки в 1,5-2,5 аллельных замещения. Любая мутация, которая позволяет продвинуться к локальному адаптивному пику, способствует следующей мутации, которая переводит генотип еще ближе к адаптивному пику.

Такой режим эволюционной динамики приводит к значительной дисперсии средней скорости эволюции в различных линиях, что согласуется с оценками, полученными на основе последовательностей.

Дж. Гиллеспи показывает, что среднее время ожидания двойной последовательной замены, в которой промежуточный аллель менее приспособлен (адаптивная долина), обратно пропорционально квадрату частоты мутаций  $u$  и обратно пропорционально численности популяции  $n$  ( $S$  и  $s$  – коэффициенты отбора на первую и вторую замену) :

$$t \sim \frac{S}{2nu^2s} .$$

Этот результат противоречит результату С. Райта [28], который показывал, что скорость пересечения адаптивных долин уменьшается с ростом размера популяции. Согласно Дж. Гиллеспи, лимитирующим фактором адаптации является именно скорость мутации, и чем больше популяция содержит различных генотипов, тем большее количество мутационных вариантов ей доступно. Согласно С. Райту, пересечение адаптивных долин осуществляется за счет генетического дрейфа, влияние которого тем сильнее, чем меньше размер популяции.

Предположения Дж. Гиллеспи о “вспышках” замен подтвердились анализом несинонимичных замен, происходящих в одном кодоне, в эволюции мыши и крысы [29] и вирусов иммунодефицита человека [30].

### 1.6 Адаптивный ландшафт М. Кимуры

В нейтральной теории молекулярной эволюции М. Кимура, Дж. Кинг (J. King) и Т. Джукс (T. Jukes) [31,32,33] утверждают, что большая часть мутаций имеют малые коэффициенты отбора, и, следовательно, их частоты в популяции главным образом определяются не отбором, а мутационно-дрейфовым равновесием.

Соответствующий адаптивный ландшафт имеет вид поверхности с незначительными пиками и долинами, такой, что переход из любой точки ландшафта в соседнюю не сопровождается большим изменением приспособленности. Важнейшей особенностью динамики на плоских ландшафтах является независимость скорости эволюции от численности популяции, при условии, что численность популяции конечна.

Плоский ландшафт М. Кимуры, соответствующий нейтральной теории, применяется в качестве нулевой модели в популяционно-генетических исследованиях. Кроме того, эволюция некоторых последовательностей, например синонимичных сайтов или межгенных участков, не несущих регуляторной функции, близка к нейтральной.

Важным обобщением нейтральной теории стала околонеутральная теория Т. Оты (T. Ohta) [34,35,36]. В этой теории вводится класс околонеутральных мутаций, имеющих малые коэффициенты отбора: слабовредные и слабополезные мутации. Вероятность их фиксации зависит от соотношения эффективности отбора и генетического дрейфа. В популяциях большой численности более эффективен отбор, а с уменьшением численности влияние отбора уменьшается, а генетического дрейфа – возрастает. Адаптивный ландшафт в этом случае представляет собой модификацию ландшафта М. Кимуры – это изломанный ландшафт с большим количеством пиков и долин небольшой амплитуды, и с небольшим количеством больших пиков и глубоких

долин. Популяция занимает на ландшафте некоторую область, а не сосредоточена в одной точке. При этом популяция малой численности может не удержаться на локальном небольшом пике в результате действия случайного дрейфа, а в популяции большой численности отбор “заметит” даже небольшой адаптивный пик.

Для околонейтральной модели эволюции на адаптивном ландшафте остается открытым вопрос, поставленный еще С. Райтом, о том, каким образом популяция большой численности сможет достичь высокого адаптивного пика, если на пути к нему располагаются адаптивные пики меньшей высоты. Получается, что чем больше эффективный размер популяции, тем труднее ей будет в этом случае оказаться на высоком адаптивном пике.

### **1.7 Адаптивные прогулки на изломанных ландшафтах**

Понимание эволюционного процесса как движения по адаптивному ландшафту, построенному в пространстве дискретных признаков (молекулярных последовательностей), привлекло в область эволюционной биологии специалистов по нелинейной динамике многокомпонентных дискретных систем. Из важнейших работ в этом направлении следует упомянуть работы С. Левина (S. Levin) и С. Кауффмана (S. Kauffman) [37,38].

С. Левин и С. Кауффман показали, что при росте размерности адаптивного ландшафта, как и предполагал Р. Фишер, локальные пики становятся седловыми точками, однако в то же время возникает множество новых локальных пиков. Таким образом, с ростом размерности ландшафт становится более изломанным и достижение глобального пика лишь в результате действия отбора становится невозможным.

Прогулка на адаптивном ландшафте (adaptive walk) – это процесс

чередования последовательности генотипов в ходе эволюции, в котором каждый следующий генотип отличается от предыдущего одной нуклеотидной заменой, а также не худшей приспособленностью. Начальная точка прогулки выбирается случайно, конечной точкой становится локальный адаптивный пик — такая точка на адаптивном ландшафте, все соседи которой имеют меньшую приспособленность. Длина адаптивной прогулки — это количество замен от начальной до конечной точки.

В работе С. Левина и С. Кауффмана адаптивная прогулка — это модель эволюции усредненного генотипа в популяции, или генотипа со средней приспособленностью. Модель содержит несколько упрощений [9, с.50]: в популяции отсутствует изменчивость, вредные мутации не фиксируются, полезные мутации не теряются, вероятность фиксации не зависит от частоты мутаций.

Несмотря на значительные и нереалистичные упрощения, модель С. Левина и С. Кауффмана, обладая значительной универсальностью, способствовала изучению динамики на изломанных ландшафтах, причем не только в эволюционной биологии, но и в других областях знания: физике, экономике, теории алгоритмов и т.д. На данный момент количество ссылок на эту работу приближается к 900.

Среди основных результатов исследования Левина-Кауффмана следует отметить несколько моментов [9, с.50]: количество локальных пиков очень велико, для генотипа длиной 100 нуклеотидов оно составляет около  $1,26 \times 10^{28}$ ; с каждым адаптивным шагом количество доступных в дальнейшем адаптивных шагов уменьшается в 2 раза; прогулки до ближайшего адаптивного пика имеют небольшую длину — в среднем 6,33 шага для генотипа в 100 нуклеотидов (следует отметить, что этот результат хорошо коррелирует с соображениями Дж. Гиллеспи о длине эволюционных “вспышек”); ожидаемое время достижения локального пика, которое включает и время поиска следующего

адаптивного шага, пропорционально длине генотипа; из любой начальной точки ландшафта можно достичь лишь небольшого количества адаптивных пиков, и, напротив, каждый адаптивный пик может быть достигнут из весьма небольшого количества начальных точек.

Пожалуй, главным результатом модели Левина-Кауффмана можно считать происходящую на изломанных ландшафтах “катастрофу сложности”: с ростом длины генотипа изломанность ландшафта возрастает, средняя высота локального пика стремится к средней высоте всех пиков на ландшафте, популяция за счет отбора не в состоянии достичь редких высоких адаптивных пиков, поэтому средняя приспособленность в популяции под действием отбора стремится к средним значениям высоты пика для всего ландшафта.

### 1.8 НК-модель С. Кауффмана

НК-модель изломанного адаптивного ландшафта С. Кауффмана [38, с.40; 39] помимо генотипов и приспособленности включает эпистатические взаимодействия: в генотипе длиной  $N$  нуклеотидов влияние локуса  $X$  на приспособленность связано еще с  $K$  локусами. В модели С. Кауффмана изломанность ландшафта регулируется соотношением параметров  $N$  и  $K$ : при  $K=0$  ландшафт гладкий, при  $K=N$  – максимально изломанный. Соответственно, длина пути до локального адаптивного пика уменьшается с ростом  $K$ .

НК-модель С. Кауффмана дает по меньшей мере два важнейших следствия. Во-первых, можно оценивать параметр  $K$ , исходя из молекулярных данных о приспособленности конкретных белков, как это сделано в работе [39]. По опытным данным для ландшафта приспособленности переменных участков антител иммунной системы человека, в котором приспособленность антитела измеряется аффинностью, т.е. силой связи антитела и антигена, удалось оценить длину адаптивной прогулки в 6-8 мутационных шагов. По длине

адаптивной прогулки и NK модели адаптивного ландшафта было вычислено значение параметра  $K=40$ . При этом длина всего варибельного участка составляет 330-360 нуклеотидов, или 110-120 аминокислот. Изучение NK-модели адаптивного ландшафта с известными параметрами позволило авторам сделать биологически важные выводы, которые нашли свое подтверждение в экспериментах: доля полученных в ходе клональной селекции антител, которые являются более приспособленными, чем исходный вариант, составляет 1-2%; сайты варибельного участка сильно различаются по степени влияния на приспособленность; из одной и той же начальной точки существует множество адаптивных путей, однако многие из них заканчиваются на одном адаптивном пике.

Во-вторых, можно сделать предположение о том, что при помощи эпистаза популяция, которая движется по адаптивному ландшафту, может в некоторой степени управлять его изломанностью, подстраивая величину  $K$ .

### 1.9 Флуктуации отбора

Несмотря на теоретические, и, с накоплением данных о последовательностях, практические успехи в изучении адаптивных ландшафтов, в большинстве работ адаптивные ландшафты рассматривались как статические объекты [10]. Однако в реальности условия среды меняются как на геологических временах (изменения климата), так и на временах десятков-сотен лет (экологические изменения). Изменения условий среды и эволюционные изменения зачастую происходят на одних и тех же временах, поскольку в условия среды для организма включаются и приспособленности его соседей по популяции, и приспособленности организмов других видов, с которыми он взаимодействует (хищников, жертв, паразитов и т.д.) [40]. Следовательно, адаптивный ландшафт сам по себе является динамическим объектом.

Следует отметить, что переход от статических адаптивных ландшафтов к динамическим в эволюционной биологии труден, поскольку традиционно математический аппарат эволюционной биологии – популяционная генетика – складывался как исследование равновесий между различными эволюционными факторами: отбором, мутациями, дрейфом, миграцией.

Исследование динамических адаптивных ландшафтов началось с изучения популяционно-генетических моделей, в которых сила отбора менялась со временем на микроэволюционных временах.

В работе М. Кимуры 1954 года [41] рассматривается динамика нейтрального в среднем аллеля при случайных флуктуациях отбора. Для этого случая показано, что в большой популяции не происходит фиксации или исчезновения аллеля в строгом смысле, однако существует сильная тенденция к состояниям, близким к фиксации или исчезновению. Это явление названо квази-фиксацией.

В работе Дж. Гиллеспи 1973 года [42] рассматривается модель эволюции в популяции бесконечного размера для двух гаплоидных генотипов, приспособленность каждого из которых меняется случайно от поколения к поколению. В итоге предпочтительным оказывается аллель с большим геометрическим средним приспособленности.

Анализ полиморфизмов в популяциях *Drosophila melanogaster* и *Drosophila simulans* и замен между этими видами при помощи чувствительного к режиму отбора популяционно-генетического теста показал, что отбор меняется на временных масштабах нейтральной эволюции [43].

### **1.10 Неравновесная динамика адаптивных ландшафтов**

В середине 1990-х годов для изучения динамических адаптивных ландшафтов был привлечен аппарат статистической физики, который позволяет

описывать динамику ансамблей популяций и потоки приспособленности [44].

На этой основе была заложена теория динамических адаптивных ландшафтов, или морских ландшафтов [45]. В этой работе рассматривается эволюционная динамика на обычном статическом ландшафте, на динамическом ландшафте, изменения которого происходят на макроэволюционных временах, и на быстро меняющемся ландшафте, в случае которого характерное время изменения ландшафта сравнимо со скоростью микроэволюционных процессов.

В случае статического ландшафта популяция достигает адаптивного пика и входит в состояние равновесия, в котором полезные замены лишь компенсируют вредные. При этом адаптации не происходит. Большинство мутаций в этом состоянии популяции являются слабовредными, поскольку популяция находится на адаптивном пике.

В случае макроэволюционного динамического адаптивного ландшафта его изменения приводят к тому, что средняя приспособленность в популяции регулярно снижается (адаптивный пик “уходит” из-под популяции). В этом случае появляются мутации, которые повышают приспособленность; они отбираются и фиксируются, и популяция движется на адаптивный пик. Этот процесс все время повторяется, в результате чего популяция не фиксируется на адаптивном пике, а находится в неравновесном стационарном состоянии: для того, чтобы возвращаться на подвижный адаптивный пик, популяции нужен положительный поток приспособленности. В результате количество полезных мутаций превышает количество вредных, однако на макроэволюционных временах повышения средней приспособленности в популяции не происходит.

Помимо теоретических результатов, работа В. Мустонена (V. Mustonen) и М. Лэссига (M. Lassig) является примером исследования эволюции молекулярных последовательностей. Для регуляторных последовательностей *Abf1* у дрожжей показан равновесный характер эволюции, а для различных классов последовательностей генома *Drosophila melanogaster* показано неравновесное состояние с положительным притоком приспособленности.

В. Мустонен и М. Лэссиг обобщили теорему Р. Фишера, включив в модель эволюции действие мутаций, дрейфа и флуктуирующего отбора и рассмотрев статистическую динамику ансамблей популяций [46]. В качестве меры адаптивной эволюции предлагается поток приспособленности, который определяется как произведение коэффициента отбора на скорость изменения частоты аллеля. Доказан рост кумулятивного потока приспособленности для большинства популяций. Лишь в малой доле популяций уменьшающейся численности происходит уменьшение потока приспособленности.

### 1.11 Дырявый ландшафт

С. Гаврилец предложил модель дырявого адаптивного ландшафта [47; 9, с.309]. Он отметил, что классический ландшафт С. Райта, с пиками и долинами, представляет собой проекцию нашего опыта жизни в трехмерном мире, в то время как реальный адаптивный ландшафт многомерен. Проблема модели С. Райта в том, что популяция не может пересечь адаптивную долину, чтобы подняться на более высокий пик приспособленности. Для решения этой проблемы С. Райт предложил свою теорию смещенного равновесия, однако математический анализ этой теории показывает, что пересечение адаптивных долин в ней возможно лишь в некоторых специальных случаях.

В модели С. Гаврильца приспособленности генотипов (которые определяются комбинациями аллелей) назначаются случайно и независимо, функция приспособленности имеет лишь 2 значения: 0 и 1.

В пространстве генотипов находится множество точек приспособленности 1, связанных друг с другом непрерывными путями. Однако встречаются “несовместимые” комбинации аллелей, приспособленность которых равна нулю, и которые представляют собой дыры в плоском адаптивном ландшафте.

Наличие таких областей, возможно состоящих из множества точек, может вызывать репродуктивную изоляцию генотипов, находящихся по разные стороны от дыры.

С помощью данной модели С. Гаврилец рассматривает эволюционную динамику видообразования под действием генетического дрейфа, мутаций, рекомбинации и миграции.

В работах С. Гаврильца рассматриваются статические ландшафты, за исключением случаев частотно-зависимого отбора.

### **1.12 Практическое применение теории адаптивных ландшафтов**

В 1970-х годах была создана теория генетических алгоритмов, которые позволяют успешно решать оптимизационные задачи в самых различных прикладных и научных областях [48,49,50]. Суть генетического алгоритма в том, чтобы путем моделирования эволюционного процесса на искусственных популяциях найти адаптивные пики в сложных пространствах, для которых аналитическое решение задачи оптимизации невозможно.

В генетическом алгоритме генотип особи кодируется строкой. Над пространством генотипов задается функция приспособленности. Алгоритм состоит из этапов начального распределения особей по адаптивному ландшафту, и далее повторяющегося цикла: мутации – скрещивание – отбор. Скрещивание происходит для двух случайно выбранных особей в популяции. Алгоритм заканчивает работу, когда улучшение максимальной приспособленности в популяции на шаг в течение некоторого количества шагов остается меньше заданного порога.

### 1.13 Изучение ландшафтов по молекулярным данным

На протяжении последних 50 лет усовершенствование лабораторных методик секвенирования [51] приводит к накоплению данных о последовательностях [52,53].

Изучение адаптивных ландшафтов по молекулярным данным возможно в нескольких направлениях.

Во-первых, возможна постановка долговременных эволюционных экспериментов, таких как эксперимент Р. Ленски (R. Lenski) [54]. Задавая определенное направление адаптации для организмов с коротким жизненным циклом (бактерий), можно измерять приспособленность раз в несколько поколений (по скорости роста колонии), и, секвенируя генотипы на промежуточных стадиях, отследить весь адаптивный путь от начального состояния до вершины адаптивного пика. Несколько параллельных экспериментов, стартующих из разных точек адаптивного ландшафта, позволяют изучить не один адаптивный путь, а некоторую часть ландшафта.

Во-вторых, возможно изучение всех отдельно взятых аминокислотных замен для определенных белков. Например, в работе [55] авторы проследили путь в 20 аминокислотных замен от встречающегося в природе белка цитохром гидроксилазы P450BM3 до пропан-монооксигеназы P450PMO. Полученный в лаборатории фермент высокоэффективен и специфичен: он способствует превращению пропана в пропиловый спирт.

В-третьих, возможно секвенирование и изучение белковой эволюции видов, приспособленных к экстремальным условиям: живущих в холодных водах или льдах, в условиях высокой температуры или давления, в условиях радиационного загрязнения и т.д. [56]. Сравнение белок-кодирующих последовательностей экстремофилов с мезофилами может дать материал для изучения соответствующих адаптивных ландшафтов.

В-четвертых, можно использовать наборы геномных данных для широкого набора видов, а также данные о внутривидовом полиморфизме. Здесь различают два подхода: метод “толстых” выравниваний, когда используются последовательности распространенного белка, скажем, цитохромоксидазы для большого количества видов (сотни и тысячи видов), и метод “длинных” выравниваний, в этом случае выравниваются полные геномы организмов разных видов, например геномы млекопитающих, позвоночных, насекомых. Такие наборы данных доступны в базе данных UCSC Genome Bioinformatics [57].

Что касается полиморфизма, то доступны данные по более чем 1000 генотипам человека [58], 200 генотипам *Drosophila melanogaster* [59] и множеству раковых генотипов [60].

Имея такие данные, можно получать информацию о частотах синонимичных и несинонимичных замен и полиморфизмов, по которой опосредованно можно судить о наличии или отсутствии положительного отбора [61,62], а также о динамике адаптивного ландшафта (см. Главу 3).

### **1.14 Адаптивные ландшафты в эволюции вирусов**

Помимо эволюции иммунного ответа, еще одной областью с большой скоростью эволюции является эволюция вирусов. В частности, для вируса ВИЧ при длине генома в 1000 нуклеотидов частота мутаций при репликации составляет  $10^{-3}$ - $10^{-4}$ ; таким образом, практически каждый новый цикл репликации порождает новые генотипы. Из одного образца крови получают множество генотипов вируса. Анализ больших наборов данных позволяет выявить особенности эволюции на адаптивном ландшафте [30].

В частности, исследователи обнаружили эффект группирования

несинонимичных замен, происходящих последовательно в аминокислотном сайте, т.н. эффект “вспышек” замен. Этот эффект более ярко выражен, чем корреляции замен в соседних сайтах, а также по сравнению с синонимичными заменами.

Для объяснения этого феномена, после исключения множества гипотез, не связанных с отбором, были предложены две модели, в которых меняется адаптивный ландшафт для данного сайта: случай “биений”, когда ландшафт меняется много раз в течение некоторого промежутка времени, оставаясь постоянным за пределами этих промежутков, и случай одноразового изменения адаптивного ландшафта. В первом случае динамика несинонимичных замен непосредственно коррелирует с динамикой ландшафта: очереди замен вызываются несколькими последовательными событиями положительного отбора. Второй случай находится в соответствии с рассмотренными ранее моделью Дж. Гиллеспи, которая предполагала наличие “вспышек”, т.е. последовательностей замен, и теорией адаптивных прогулок Кауффмана-Левина.

Различие между этими двумя режимами динамики адаптивного ландшафта – дело будущего исследования, для которого потребуются данные о более длинных адаптивных прогулках.

### **1.15 Выводы**

Адаптивный ландшафт служит важнейшей концепцией эволюционной биологии, важнейшим методологическим приемом исследования эволюционных процессов. Процесс эволюции рассматривается как движение по адаптивному ландшафту, что позволяет строить модели такого движения, сравнивать молекулярные данные о частотах замен в эволюции с модельными предсказаниями, и таким образом верифицировать утверждения, заложенные в

основу моделей.

Адаптивный ландшафт меняется в процессе эволюции в глобальном масштабе за счет изменений факторов среды, так и в масштабе одного локуса, за счет эпистатических взаимодействий с другими сайтами. Раскрыть механизмы этих изменений, с учетом основных факторов эволюции, на основе теоретических моделей и молекулярных данных — задача будущих исследований. В частности, в настоящей работе исследуется динамика однолокусного адаптивного ландшафта после аминокислотной замены.

## 2 Динамика триаллельного адаптивного ландшафта

Экспериментальное изучение динамики адаптивного ландшафта затруднено. Однако его изменения могут приводить к паттернам во внутривидовом полиморфизме или межвидовой дивергенции, которые можно наблюдать методами сравнительной геномики, и такие данные позволяют сопоставлять различные модели динамики адаптивного ландшафта.

### 2.1 Прямые, обратные и боковые замены в триаллельном сайте

Пусть в некотором сайте белок-кодирующего гена в эволюционной истории вида произошла замена  $A \rightarrow B$ . Со временем в этом же сайте может произойти обратная замена (реверсия)  $B \rightarrow A$ , боковая замена  $B \rightarrow C$ :  $C \neq A, B$ , либо сайт может остаться инвариантным.

Прямая замена  $A \rightarrow B$  может быть вредной ( $w(A) \gg w(B)$ ), околонеutralной ( $w(A) \sim w(B)$ ) или полезной ( $w(A) \ll w(B)$ ). Аллель, занимающий сайт, имеет, как правило, достаточно высокую приспособленность (не сильно ниже, чем приспособленность его мутационных соседей), поскольку он является продуктом предшествующей истории адаптаций. Напротив, приспособленность аллелей, не наблюдаемых в сайте, может быть произвольной. Возникает вопрос о динамике отношения приспособленностей аллелей  $A$  и  $B$  со временем, прошедшим после прямой замены.

Рассмотрим случай околонеutralной (слабополезной) прямой замены  $A \rightarrow B$  (Рисунок 2.1). После такой замены возможно увеличение  $w(B)$  и/или уменьшение  $w(A)$ .

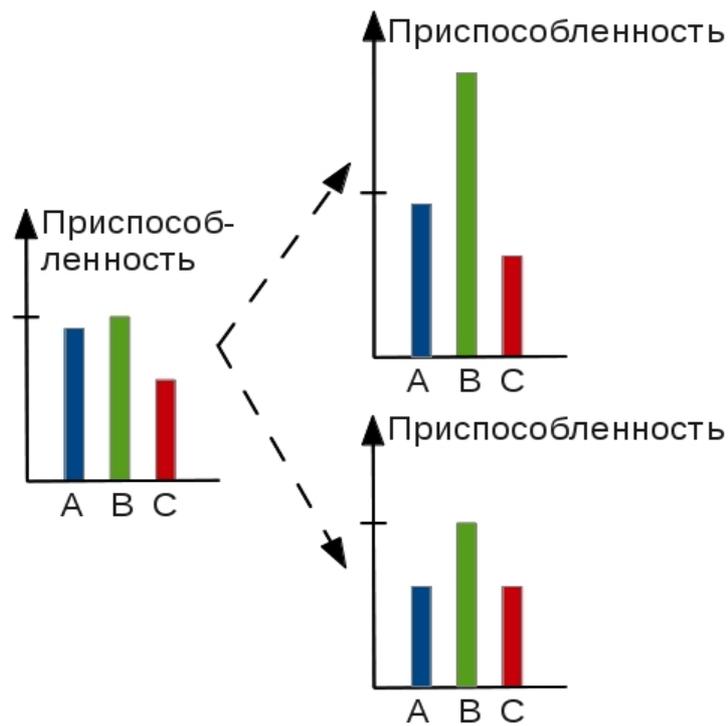


Рисунок 2.1. Два возможных типа динамики адаптивного ландшафта в случае околонеutralной замены

Пусть аллели  $A$ ,  $B$ ,  $C$  имеют относительные приспособленности  $w(A) = 1 - s_1(t)$ ,  $w(B) = 1$  и  $w(C) = 1 - s_2(t)$ , где  $t$  соответствует времени после исходной замены  $A \rightarrow B$ , а  $s_1, s_2 > 0$  – коэффициенты отбора. Тогда частоты слабовредных замен  $f(B \rightarrow A)$ ,  $f(B \rightarrow C)$  относительно нейтральной замены определяются уравнением 3.14 из [33]:

$$f(B \rightarrow A) = \frac{S_1}{e^{S_1} - 1}$$

$$f(B \rightarrow C) = \frac{S_2}{e^{S_2} - 1},$$

где  $S_1 = 4N_e s_1$ ,  $S_2 = 4N_e s_2$ , и  $N_e$  – эффективная численность популяции. Необходимо отметить, что у Кимуры коэффициент отбора для вредной замены (каковой является замена  $B \rightarrow A$ ) отрицательный, а согласно нашим определениям (и, например, определениям Дж. Гиллеспи), он положительный.

## 2.2 Случай увеличения приспособленности производного аллеля

Предположим, что после замены  $A \rightarrow B$  абсолютные значения приспособленности  $w_a(A)$  и  $w_a(C)$  остаются постоянными, а  $w_a(B)$  возрастает со временем (Рисунок 2.1). Это соответствует уменьшению относительной приспособленности  $w(A)$  и  $w(C)$ , то есть увеличению коэффициентов отбора  $s_1$  и  $s_2$ :

$$s_1 = s_1(0) + g(t); \quad g(t) > 0; \quad dg/dt > 0$$

$$s_2 = s_2(0) + g(t)$$

$$s_1(0) < s_2(0) \Leftrightarrow w(A, 0) > w(C, 0)$$

Отношение  $R(t)$  частот замен  $B \rightarrow A$  и  $B \rightarrow C$  будет равно

$$R(t) = \frac{f(B \rightarrow A)}{f(B \rightarrow C)} = \frac{\frac{S_1}{e^{S_1} - 1}}{\frac{S_2}{e^{S_2} - 1}} = \frac{4N_e(s_1(0) + g(t))}{4N_e(s_2(0) + g(t))} \cdot \frac{e^{4N_e(s_2(0) + g(t))} - 1}{e^{4N_e(s_1(0) + g(t))} - 1}.$$

Численное решение показывает, что  $R(t)$  возрастает со временем (Рисунок 2.2).

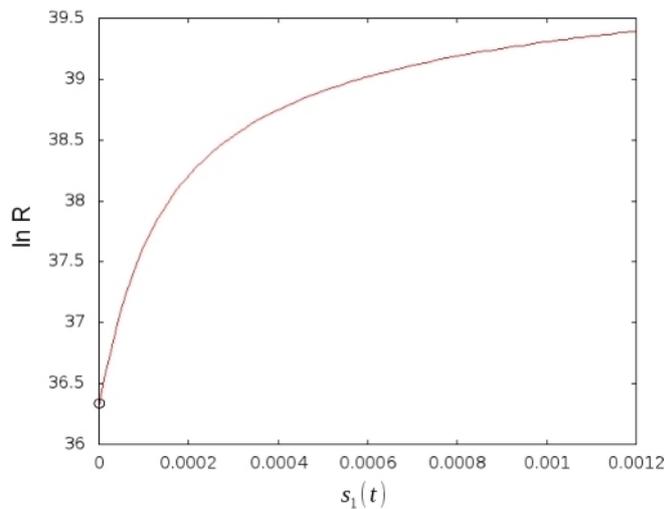


Рисунок 2.2. Отношение частот замен  $A \rightarrow B$  и  $A \rightarrow C$  как функция силы отбора против замены  $B \rightarrow A$ , в предположении, что приспособленность аллеля  $B$  увеличивается, а приспособленности аллелей  $A$  и  $C$  остаются постоянными.

$$N_e = 10\,000; \quad s_2 - s_1 = 0,001$$

Докажем возрастание  $R(t)$  аналитически.

Введем новую функцию  $x(t)$  и коэффициент  $a$ :

$$x(t) = s_1(0) + g(t); \frac{dx}{dt} > 0$$

$$a = s_2(0) - s_1(0),$$

тогда  $s_2(0) + g(t) = a + s_1(0) + g(t) = a + x(t)$ .

Выразим  $R$  через  $x(t)$  и  $a$ :

$$R(x) = \frac{x}{x+a} \cdot \frac{e^{4N_e(x+a)} - 1}{e^{4N_e x} - 1} = \frac{C_1(x)}{C_1(x) + C_2} \cdot \frac{e^{C_1(x) + C_2} - 1}{e^{C_1(x)} - 1}; C_1(x) = 4N_e x; C_1(x) > 1; C_2 = 4N_e a; C_2 > 0$$

Для того, чтобы определить, возрастает  $R(x)$  или убывает, исследуем знак производной  $dR/dx$ :

$$\frac{dR}{dx} = \frac{4N_e x^2 (e^{4N_e x} - e^{4N_e x + 4N_e a}) + 4aN_e x (e^{4N_e x} - e^{4N_e a + 4N_e x}) + a - ae^{4N_e x} + ae^{4N_e a + 8N_e x} - ae^{4N_e a + 4N_e x}}{(x+a)^2 (e^{4N_e x} - 1)^2}$$

$$\frac{dR}{dx} = \frac{4N_e x (x+a) (e^{4N_e x} - e^{4N_e x + 4N_e a}) + a(1 - e^{4N_e x}) (1 - e^{4N_e a + 4N_e x})}{(x+a)^2 (e^{4N_e x} - 1)^2}$$

$$\frac{dR}{dx} = \frac{4N_e x (x+a) e^{4N_e x} (e^{-2N_e a} - e^{2N_e a}) + a(1 - e^{4N_e x}) (e^{-2N_e a} - e^{2N_e a + 4N_e x})}{(x+a)^2 (e^{4N_e x} - 1)^2 e^{-2N_e a}}$$

$$\frac{dR}{dx} = \frac{-4N_e x (x+a) (e^{2N_e a} - e^{-2N_e a}) + a(e^{-2N_e x} - e^{2N_e x}) (e^{-2N_e a - 2N_e x} - e^{2N_e a + 2N_e x})}{(x+a)^2 (e^{4N_e x} - 1)^2 e^{-2N_e a} e^{-4N_e x}}$$

$$\frac{dR}{dx} = \frac{-8N_e x (x+a) \sinh(2N_e a) + 4a \sinh(2N_e x) \sinh(2N_e a + 2N_e x)}{(x+a)^2 (e^{4N_e x} - 1)^2 e^{-2N_e a} e^{-4N_e x}}$$

$$\sinh(2N_e x) \sinh(2N_e a + 2N_e x) = \frac{\cosh(2N_e x + 2N_e a + 2N_e x) - \cosh(2N_e x - 2N_e a - 2N_e x)}{2} =$$

$$= \frac{\cosh(4N_e x + 2N_e a) - \cosh(2N_e a)}{2}$$

$$\frac{dR}{dx} = \frac{-8N_e x (x+a) \sinh(2N_e a) + 2a \cosh(2N_e a + 4N_e x) - 2a \cosh(2N_e a)}{(x+a)^2 (e^{4N_e x} - 1)^2 e^{-2N_e a} e^{-4N_e x}}.$$

Знак производной определяется выражением

$$C_1(x) = -8N_e x (x+a) \sinh(2N_e a) + 2a \cosh(2N_e a + 4N_e x) - 2a \cosh(2N_e a).$$

Рассмотрим неравенство

$$C_1(x) = -4N_e x(x+a) \sinh(2N_e a) + a \cosh(2N_e a + 4N_e x) - a \cosh(2N_e a) > 0$$

$$a \cosh(2N_e a + 4N_e x) > 4N_e x(x+a) \sinh(2N_e a) + a \cosh(2N_e a) .$$

Из разложения левой части неравенства  $a \cosh(2N_e a + 4N_e x)$  в ряд Тейлора в окрестности точки  $x=0$  верно неравенство

$$a \cosh(2N_e a + 4N_e x) = a \cosh(2N_e a) + 4N_e x a \sinh(2N_e a) + 8N_e^2 x^2 a \cosh(2N_e a) + \frac{32}{3} N_e^3 x^3 a \sinh(2N_e a) + b(x) > 4N_e x(x+a) \sinh(2N_e a) + a \cosh(2N_e a)$$

$$a \cosh(2N_e a) + 4N_e x a \sinh(2N_e a) + 8N_e^2 x^2 a \cosh(2N_e a) + \frac{32}{3} N_e^3 x^3 a \sinh(2N_e a) \geq 4N_e x(x+a) \sinh(2N_e a) + a \cosh(2N_e a)$$

$$8N_e^2 x^2 a \cosh(2N_e a) + \frac{32}{3} N_e^3 x^3 a \sinh(2N_e a) \geq 4N_e x^2 \sinh(2N_e a)$$

$$2N_e a \cosh(2N_e a) + \frac{8}{3} N_e^2 x a \sinh(2N_e a) \geq \sinh(2N_e a)$$

$$2N_e a \cosh(2N_e a) + \frac{8}{3} N_e^2 x a \sinh(2N_e a) - \sinh(2N_e a) \geq 0 .$$

Рассмотрим функцию

$$z(x) = 2N_e a \cosh(2N_e a) + \frac{8}{3} N_e^2 x a \sinh(2N_e a) - \sinh(2N_e a)$$

$$z(0) = 0 .$$

Рассмотрим производную  $dz/dt$

$$\frac{dz(x)}{dx} = \frac{8}{3} N_e^2 a \sinh(2N_e a) > 0 .$$

$z(x)$  монотонно возрастает и  $z(0) = 0$ , таким образом

$$z(x) \geq 0 \Rightarrow C_1(x) > 0 \Rightarrow \frac{dR}{dx} > 0 \Rightarrow \frac{dR}{dt} > 0 .$$

В рассмотренном случае роста приспособленности аллеля В отношение частот (вероятностей) обратных замен  $B \rightarrow A$  и  $B \rightarrow C$  после прямой замены ( $A \rightarrow B$ ) увеличивается со временем.

### 2.3 Случай уменьшения приспособленности предкового аллеля

Предположим, что после исходной замены  $A \rightarrow B$  абсолютная приспособленность  $w_a(A)$  уменьшается с течением времени, а абсолютные приспособленности  $w_a(B)$  и  $w_a(C)$  остаются неизменными. Это соответствует уменьшению относительной приспособленности  $w(A)$ , т.е. коэффициент отбора  $s_1$  увеличивается, а  $s_2$  — остается постоянным:

$$s_1 = s_1(0) + g(t); s_1 > 0; g(t) > 0; dg/dt > 0$$

$$s_2 = s_2(0); s_2 > 0$$

$$s_1(0) < s_2(0) \Leftrightarrow w(A,0) > w(C,0)$$

$$S_1 = 4N_e s_1$$

$$S_2 = 4N_e s_2$$

$$R(t) = \frac{f(B \rightarrow A)}{f(B \rightarrow C)} = \frac{\frac{S_1}{e^{S_1} - 1}}{\frac{S_2}{e^{S_2} - 1}} = \frac{e^{4N_e s_2(0)} - 1}{4N_e s_2(0)} \cdot \frac{4N_e (s_1(0) + g(t))}{e^{4N_e (s_1(0) + g(t))} - 1} = C_4 \cdot \frac{4N_e (s_1(0) + g(t))}{e^{4N_e (s_1(0) + g(t))} - 1}, C_4 > 0.$$

Численный расчет показывает, что  $R(t)$  уменьшается со временем (Рисунок 2.3).

Докажем это аналитически. Введем новую функцию  $x(t)$ :

$$x(t) = s_1(0) + g(t); \frac{dx}{dt} > 0$$

Тогда  $R(t)$  примет вид

$$R(t) = C_4 \cdot \frac{4N_e x(t)}{e^{4N_e x(t)} - 1}, C_4 > 0$$

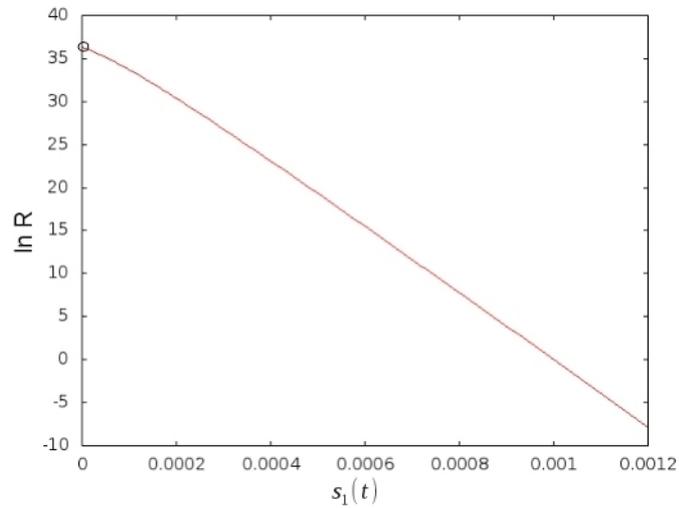


Рисунок 2.3. Отношение частот замен  $A \rightarrow B$  и  $A \rightarrow C$  как функция силы отбора против замены  $B \rightarrow A$ , в предположении, что приспособленность аллеля  $A$  уменьшается, а приспособленности аллелей  $B$  и  $C$  остаются постоянными.

$$N_e = 10,000; \quad s_2(0) = 0,001$$

Исследуем знак производной  $dR/dt$ :

$$\frac{dR}{dt} = C_4 \cdot \frac{(e^{4N_e x(t)} - 1) \cdot 4N_e \frac{dx}{dt} - 4N_e x(t) e^{4N_e x(t)} 4N_e \frac{dx}{dt}}{x^2(t)}$$

$$\frac{dR}{dt} = 4N_e \frac{dx}{dt} C_4 \cdot \frac{e^{4N_e x(t)} - 1 - 4N_e x(t) e^{4N_e x(t)}}{x^2(t)} = 4N_e \frac{dx}{dt} C_4 \cdot \frac{e^{4N_e x(t)} (1 - 4N_e x(t)) - 1}{x^2(t)}$$

$$s_1(0) > \frac{1}{4N_e} \Rightarrow 4N_e x(t) > 1 \Rightarrow \frac{dR}{dt} < 0$$

Таким образом, в рассмотренном случае уменьшения приспособленности аллеля  $A$  отношение частот (вероятностей) замен  $B \rightarrow A$  и  $B \rightarrow C$  уменьшается монотонно со временем.

## 2.4 Выводы

Проведенный анализ показывает, что при увеличении приспособленности аллеля В и при уменьшении приспособленности аллеля А динамика отношения частот обратных и боковых замен имеет противоположный характер. Это позволяет различить эти два сценария, используя молекулярные данные (см. Главу 3). Назовем изучение динамики отношения частот обратных и боковых замен АВС-тестом.

Помимо описанных сценариев временной динамики адаптивного ландшафта на динамику частот замен может влиять пространственная неоднородность адаптивного ландшафта (вдоль белковой последовательности). Этот эффект изучается в Главе 4.

### 3 Анализ динамики адаптивного ландшафта в эволюции позвоночных и насекомых

#### 3.1 Филогенетический анализ динамики адаптивных ландшафтов

Для тестирования гипотез о динамике адаптивных ландшафтов были использованы методы сравнительной геномики, эволюционные события были реконструированы с помощью сопоставления геномов нынеживущих видов. Были восстановлены сложные эволюционные сценарии, при которых в одной эволюционирующей линии происходило более одного события; для этого были необходимы достаточно большие филогении.

Для анализа были выбраны те геномы из доступных в базе UCSC Genome Browser [57], которые:

- имеют хорошее качество секвенирования (высокое покрытие);
- филогенетическое дерево которых можно поделить на сегменты сравнимой длины (Рисунок 3.1);
- топология филогенетического дерева которых позволяет восстановить событие реверсии методом наибольшей экономии [63] (Рисунок 3.2);
- полная высота филогенетического дерева составляет порядка 1 Ks (синонимичной замены на синонимичный сайт).

По этим критериям были отобраны два независимых набора данных (Рисунок 3.1):

- 9 геномов позвоночных: человека *Homo sapiens*, домовый мыши *Mus musculus*, собаки-боксеры *Canis lupus familiaris*, саванного слона *Loxodonta africana*, домового опоссума *Monodelphis domestica*, уткуноса *Ornithorhynchus anatinus*, банкивской джунглевой курицы *Gallus gallus*, когтистой шпорцевой лягушки *Xenopus tropicalis* и рыбы *Danio rerio*;
- 8 геномов насекомых: плодовых мушек *Drosophila melanogaster*, *Drosophila sechellia*, *Drosophila yakuba*, *Drosophila ananassae*, *Drosophila*

*pseudoobscura*, *Drosophila virilis*, малярийного комара *Anopheles gambiae* и хрущака малого булавоусого *Tribolium castaneum*.

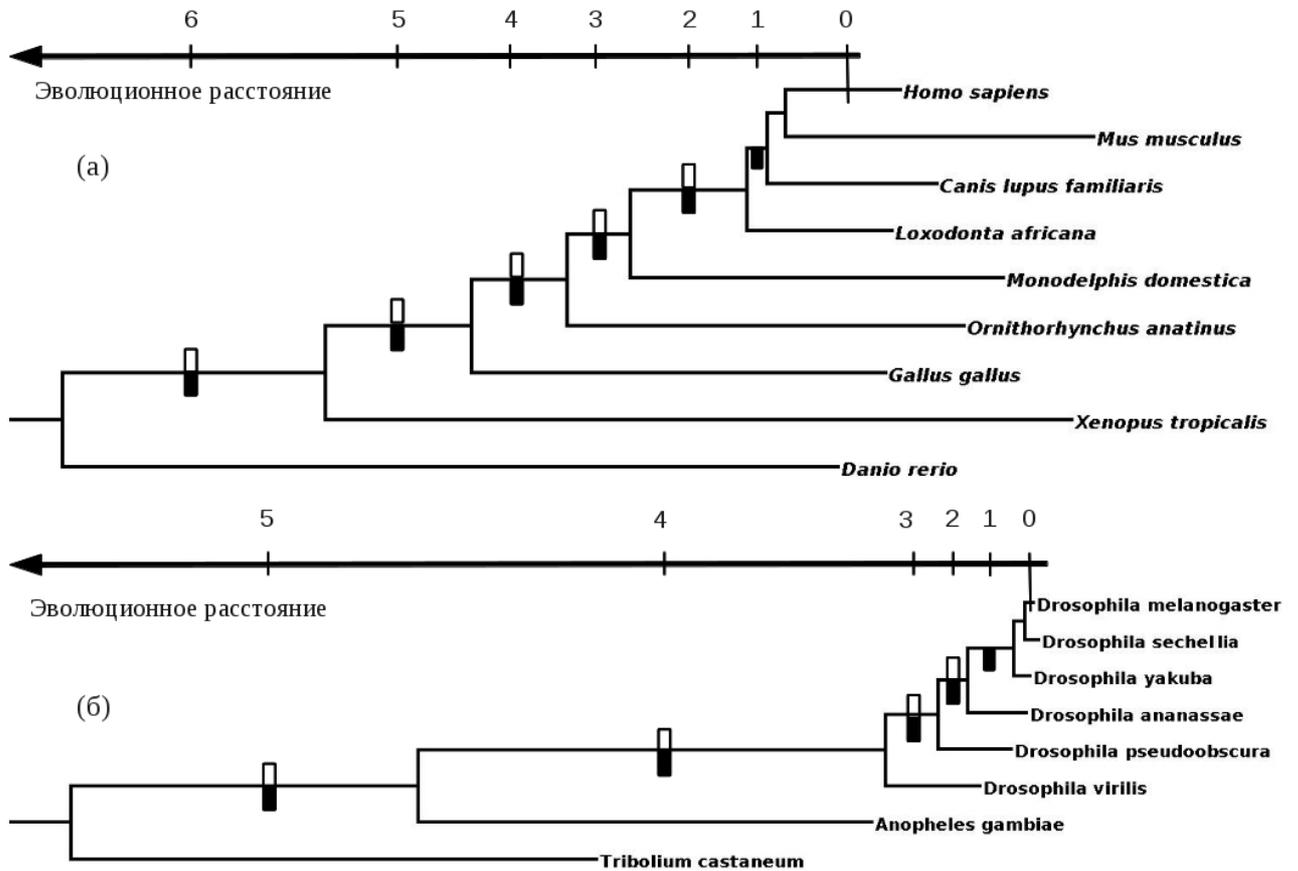


Рисунок 3.1. Филогенетические деревья анализируемых видов: позвоночных (а) и насекомых (б). Топологии филогенетических деревьев построены согласно справочным филогенетическим деревьям позвоночных и насекомых на сайте UCSC. По горизонтальной оси отложены сегменты, которые соответствуют эволюционным расстояниям между заменами

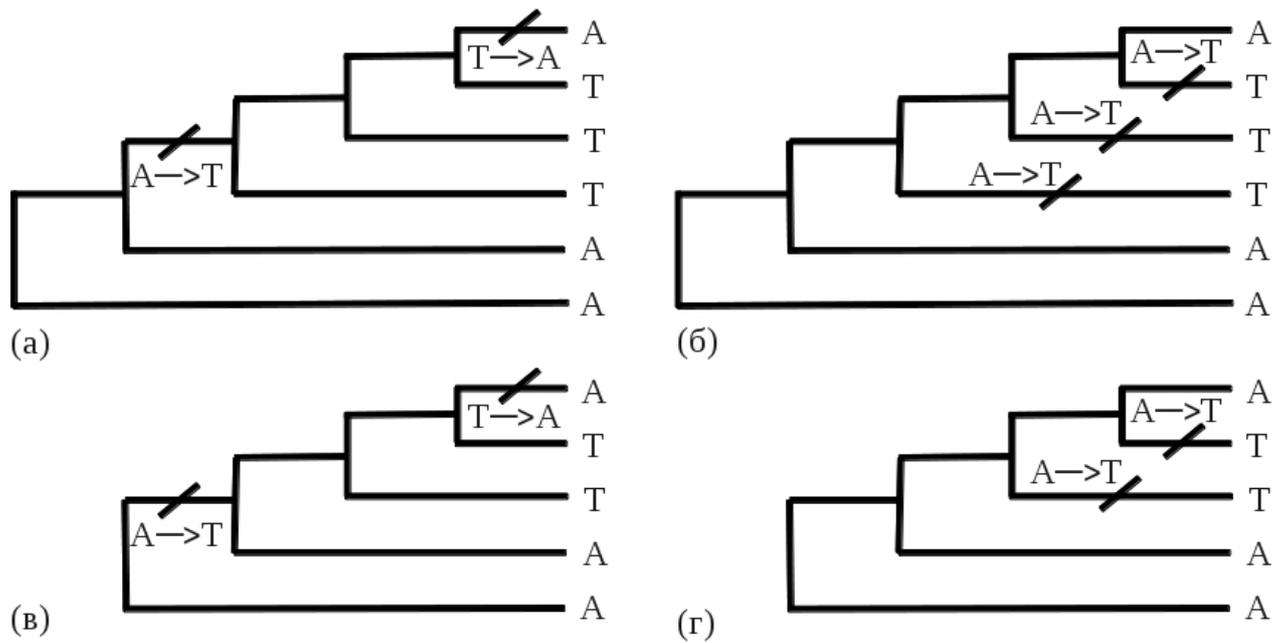


Рисунок 3.2. Восстановление событий прямой и обратной замены методом наибольшей экономии эволюционных событий. Справа показаны примеры наблюдаемых аминокислот в каждом из видов, слева — восстановленная по ним последовательность аминокислотных замен на филогении

Для наблюдаемого состояния сайта для филогенетического дерева шести видов (АТТТАА) на Рисунке 3.2 показаны 2 альтернативных сценария эволюции: (а) – требует двух замен, (б) – трех. В соответствии с методом наибольшей экономии выбирается эволюционный сценарий с наименьшим количеством замен. В случае топологии филогенетического дерева на Рисунке 3.2а-б положения всех событий могут быть зафиксированы однозначно. В случае другой топологии (Рисунок 3.2в-г) реверсия  $A \rightarrow B \rightarrow A$  равновероятна сценарию с двумя заменами  $A \rightarrow B$ ; в этом случае метод наибольшей экономии не позволяет однозначно установить филогенетические положения замен.

Для восстановления эволюционных событий был также использован метод наибольшего правдоподобия, который оценивает правдоподобие множества эволюционных сценариев при заданном филогенетическом дереве, модели

эволюции и конечных эволюционных состояниях [63].

Филогенетический анализ позволяет установить положение замены с точностью до сегмента филогенетического дерева. Поскольку середина сегмента является точкой, среднее расстояние до которой от всех замен, попавших в данный сегмент, является минимальным, будем полагать, что все замены происходят в серединах сегментов.

Будем считать, что в сегменте филогенетического дерева в данном сайте произошла замена, если нуклеотиды во втором положении кодона на концах сегмента различны. Под заменой А→В будем понимать аминокислотную замену, которая произошла в результате одной нуклеотидной замены во втором положении триплета.

Для изучения молекулярной эволюции белок-кодирующих последовательностей необходимо построить множественные выравнивания последовательностей из разных видов, имеющих общее происхождение – ортологичных последовательностей [64].

Базы ортологичных генов были подготовлены на основании полногеномных выравниваний, доступных в базе UCSC Genome Browser [57]. Основные шаги протокола подготовки данных для позвоночных представлены в Таблице 3.1.

Таблица 3.1. Подготовка данных

№	Описание
1	Загрузка множественных множественных выравниваний всех позвоночных для всех хромосом человека. <a href="http://hgdownload.cse.ucsc.edu/goldenPath/hg19/multiz46way/maf/chr1.maf.gz">http://hgdownload.cse.ucsc.edu/goldenPath/hg19/multiz46way/maf/chr1.maf.gz</a>
2	Загрузка последовательности генома человека (для всех хромосом). <a href="http://hgdownload.cse.ucsc.edu/goldenPath/hg19/chromosomes/chr1.fa.gz">http://hgdownload.cse.ucsc.edu/goldenPath/hg19/chromosomes/chr1.fa.gz</a>
3	Извлечение нужного подмножества видов maf_order chr1.maf hg19 mm9 canFam2 loxAfr3 monDom5 ornAna1 galGal3 xenTro2 danRer6 > chr1.mysps.maf
4	Перекодировка из формата maf в формат mfa lomem-maf2fasta chr1.fa chr1.mysps.maf fasta@ iupac2n > chr1.mysps.mfa
5	Получение координат экзонов из файлов аннотации <a href="http://hgdownload.cse.ucsc.edu/goldenPath/hg19/database/knownCanonical.txt.gz">http://hgdownload.cse.ucsc.edu/goldenPath/hg19/database/knownCanonical.txt.gz</a> <a href="http://hgdownload.cse.ucsc.edu/goldenPath/hg19/database/knownGene.txt.gz">http://hgdownload.cse.ucsc.edu/goldenPath/hg19/database/knownGene.txt.gz</a> get_exons_fromGBannot_all_in_line_v2.pl
6	Пересчет координат экзонов из формата аннотаций в формат выравнивания annot2align_coord.pl chr1.exons 1 > chr1.a.exons
7	Извлечение множественных выравниваний экзонов нужных видов cds_extract_vert_v8.pl chr1.a.exons 1

Данные для насекомых были подготовлены аналогичным образом.

Характеристики полученных наборов данных для позвоночных и насекомых представлены в Таблице 3.2.

Таблица 3.2. Количество проанализированных генов и кодонов

	Позвоночные	Насекомые
Видов	9	8
Генов	7967	8477
Кодонов	10 441 107	8 838 651

Также использовались данные о внутривидовом полиморфизме в популяциях *Homo sapiens* (1094 особи) и *Drosophila melanogaster* (162 особи). Данные о полиморфизме *Homo sapiens* были загружены с сайта проекта “1000 геномов” [58], данные о полиморфизме *Drosophila melanogaster* – с сайта проекта “Drosophila genetic reference panel” [59].

Общая схема расчётов приведена на Рисунке 3.3.

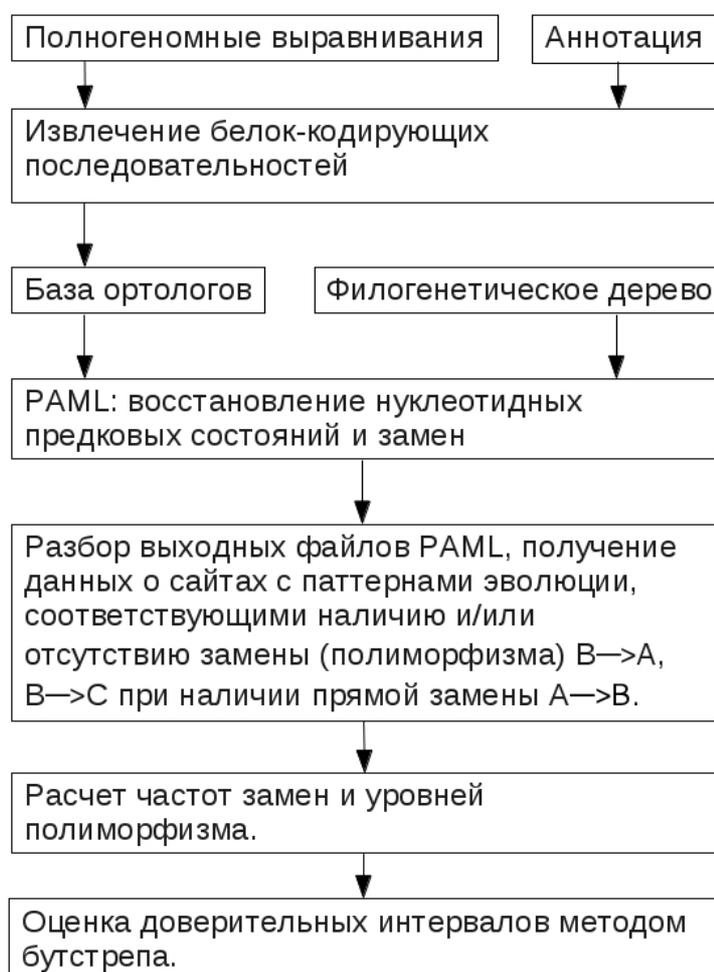


Рисунок 3.3. Схема проведения расчётов

Результаты расчетов частот замен не зависели от метода восстановления эволюционных событий – наибольшей экономии или наибольшего правдоподобия. Метод наибольшего правдоподобия позволил получить больше временных точек для анализа динамики адаптивного ландшафта, так как он не

имеет таких топологических ограничений на структуру дерева, как метод наибольшей экономии.

Поэтому состояния во внутренних узлах филогении были восстановлены методом наибольшего правдоподобия на основе имеющейся филогении и состояний в терминальных ветках дерева. Использовалась утилита `codeml` из пакета PAML [65], которая реализует модель кодонной эволюции Н. Голдмана (N. Goldman) и Дж. Янга (Z. Yang) [66].

По горизонтальной оси на Рисунке 3.1 отложен эволюционный возраст аллеля В, который возник в результате замены А→В, в предположении, что замены происходят в середине сегмента. Прямоугольниками отмечены сегменты, в которых могла произойти прямая замена А→В. Черными прямоугольниками под прямой отмечены те сегменты, для прямых замен в которых проводился анализ полиморфизма в популяции *Homo sapiens* или *Drosophila melanogaster*; белыми прямоугольниками отмечены те сегменты, для прямых замен в которых проводился анализ дивергенции в терминальной эволюционной ветви: между линиями человека и общего предка *Homo sapiens* и *Mus musculus*, или *Drosophila melanogaster* и общего предка *Drosophila melanogaster* и *Drosophila sechellia*. В случае полиморфизма были проанализированы 6(5) эволюционных расстояний для дерева позвоночных (насекомых), а в случае дивергенции – 5(4); поскольку реверсия замен, произошедших на сегменте, соответствующем эволюционному расстоянию 1, не могла быть установлена однозначно, прямые замены на этом сегменте исключались из анализа дивергенции.

В сайтах, где произошла прямая замена А→В, измерялся уровень полиморфизма в популяции *Homo sapiens* (*Drosophila melanogaster*), а также частоты замен на линии *H. sapiens* после её расхождения с линией *Mus musculus* (на линии *D. melanogaster* после её расхождения с *D. sechellia*).

Измерение динамики уровня полиморфизма и частот обратных замен

позволяет установить действие очищающего (отрицательного) отбора. Наряду с двумя независимыми наборами данных, две измеряемые статистики — уровень полиморфизма и частота обратных замен — увеличивают количество повторностей и значимость получаемой зависимости.

### 3.2 Динамика частот обратных и боковых замен

Для позвоночных и насекомых мы изучили зависимость уровня полиморфизма и скорости замен от эволюционного возраста аминокислоты, присутствующей в современном виде (Рисунок 3.1). Для сайта, в котором произошла прямая аминокислотная замена  $A \rightarrow B$  в некоторый момент времени в прошлом, мы сравнили данные по обратным  $B \rightarrow A$  и боковым  $B \rightarrow C$  полиморфизмам (или заменам), где  $C$  — это одна из боковых аминокислот, т.е. отличных от  $A$  и  $B$ .

Частота ревертирующих полиморфизмов  $B \rightarrow A$  может меняться с возрастом  $B$  в соответствии с изменением приспособленности  $A$  или  $B$ . Однако боковые аминокислоты  $C$  отсутствовали в сайте, насколько мы можем судить по имеющимся данным; поэтому замены  $A \rightarrow B$  и последующая эволюция не должны влиять на приспособленность аминокислот  $C$ . Значит, частота замен  $B \rightarrow C$  должна меняться только лишь в силу изменения приспособленности  $B$ .

Данные по полиморфизмам (Рисунок 3.4а-в) свидетельствуют о том, что внутрипопуляционное предпочтение предкового аллеля  $A$ , полученного в результате обратной мутации, уменьшается с возрастом текущего фиксированного аллеля  $B$ . Эта же величина для боковых аллелей  $C$  либо независима от возраста  $B$ , либо уменьшается с этим возрастом (в *Drosophila*, Рисунок 3.4б). Однако предпочтение предкового аллеля  $A$  уменьшается сильнее, чем боковых аллелей  $C$ . В результате отношение частот

полиморфизмов  $B \rightarrow A$  к  $B \rightarrow C$  падает в  $\sim 3$  раза за то время, пока возраст  $B$  увеличивается с очень молодого до такого, что  $\sim 1$  нуклеотидная замена произошла в синонимичном сайте с момента замены  $A \rightarrow B$  в сайте  $x$ . Данные по аминокислотным заменам дают сходные зависимости (Рисунок 3.4 г-е).

Наблюдаемые зависимости сохранялись качественно, когда рассматривались только те аминокислотные замены  $A \rightarrow B$ , которым соответствовали нуклеотидные замены слабого основания (AT) на сильное (CG), что исключает влияние эффекта смещенной геной конверсии (Рисунок 3.5, 3.6).

Наблюдаемые зависимости также сохранялись, если замены  $A \rightarrow B$  были биохимически консервативными или, напротив, радикальными (Рисунок 3.7, 3.8). Радикальность аминокислотной замены мерялась как расстояние между аминокислотами по шкале Мияты [67]. В случае консервативных замен  $A \rightarrow B$  реверсии были более частыми, а боковые полиморфизмы или замены были менее частыми; однако во всех случаях отношение частот замен  $B \rightarrow A/B \rightarrow C$  уменьшалось со временем как для полиморфизмов, так и для межвидовых отличий.

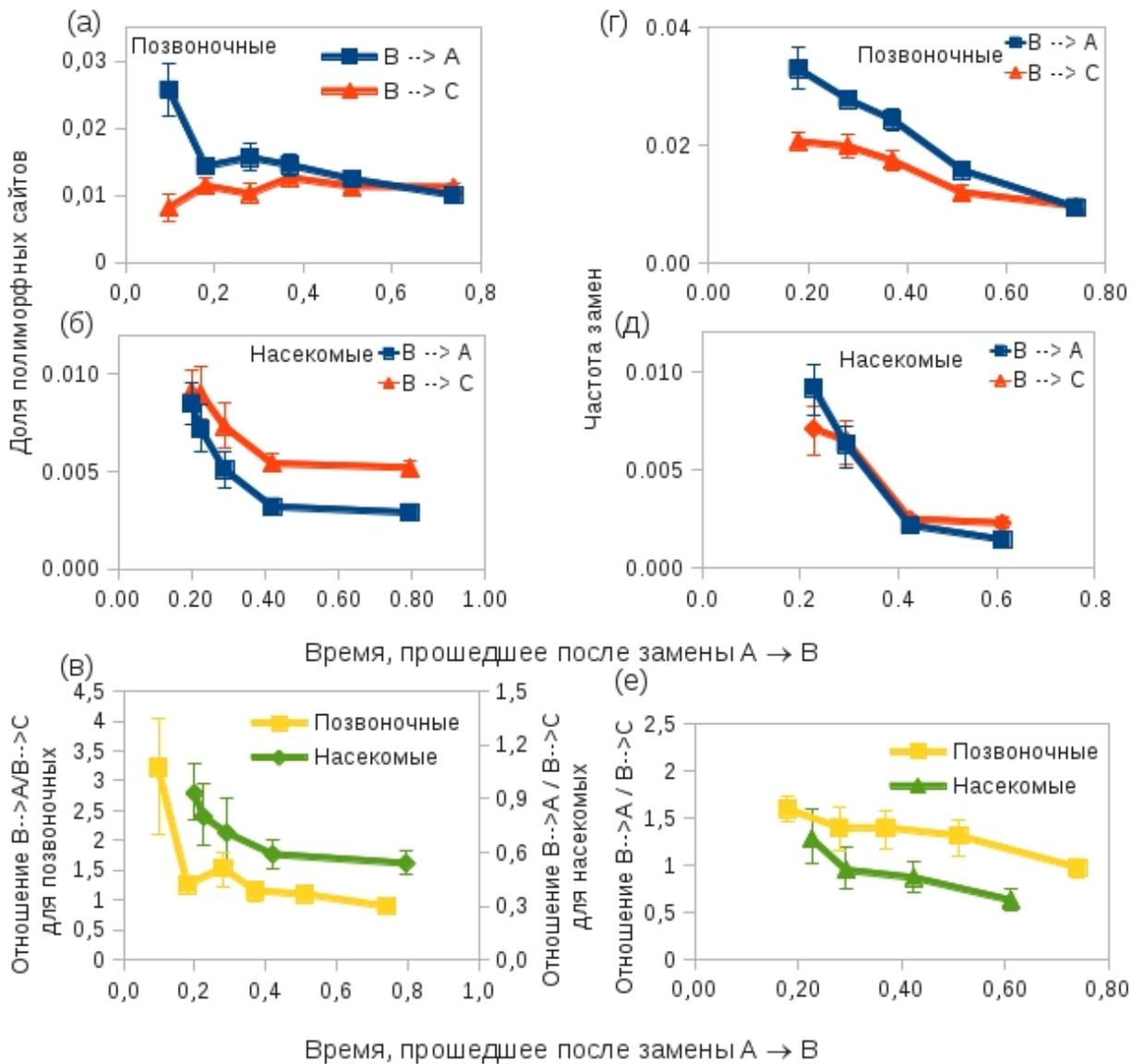


Рисунок 3.4. Зависимость уровня полиморфизма (а-в) и частот замен (г-е) от возраста предпочтительного в данный момент времени аллеля В, который зафиксировался ранее в результате замены  $A \rightarrow B$ . По оси абсцисс отложено время, прошедшее после замены  $A \rightarrow B$ , измеренное в количестве нуклеотидных замен на четырехкратно вырожденный синонимический сайт (а,б,г,д). Доля сайтов среди сайтов с заменой  $A \rightarrow B$ , которые несут аллель А (синяя линия) или один из двух аллелей С (красная линия) (а,б), или в которых произошла замена  $B \rightarrow A$  (синяя линия) или  $B \rightarrow C$  (красная линия) на терминальной ветви (г,д). (в,е) отношение количества полиморфизмов  $B \rightarrow A$  к  $B \rightarrow C$  (в), и замен (е): позвоночные (желтая линия), насекомые (зеленая линия). (а,г) – позвоночные, (б,д) – насекомые. 95% доверительные интервалы были определены методом бутстрапа (10 000 выборок)

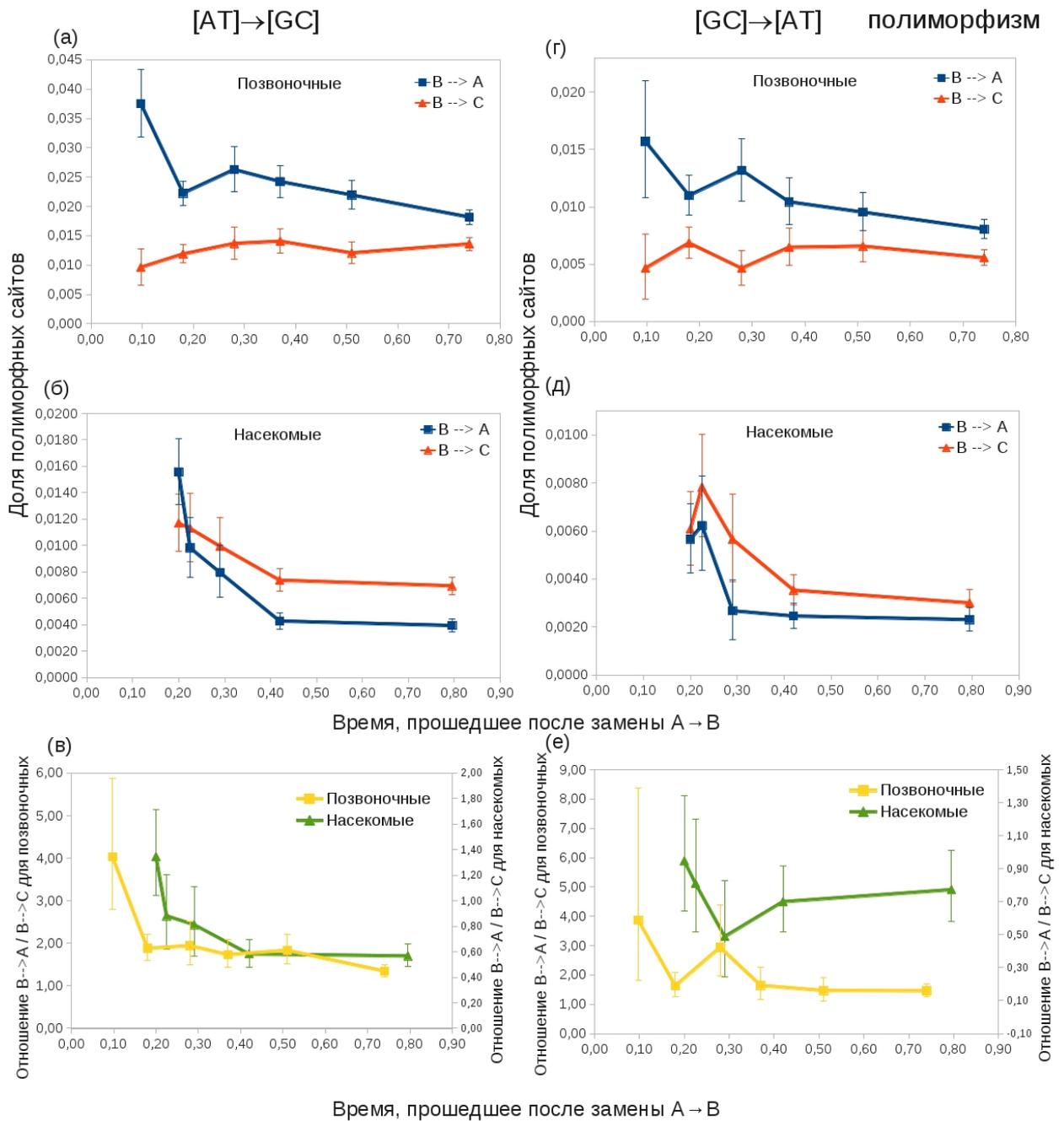


Рисунок 3.5. Зависимость уровня полиморфизма от возраста предпочтительного в данный момент времени аллеля В, который зафиксировался ранее в результате замены А→В. а,б,в: замена АТ→ГС во втором положении кодона; г,д,е: замена ГС→АТ во втором положении кодона

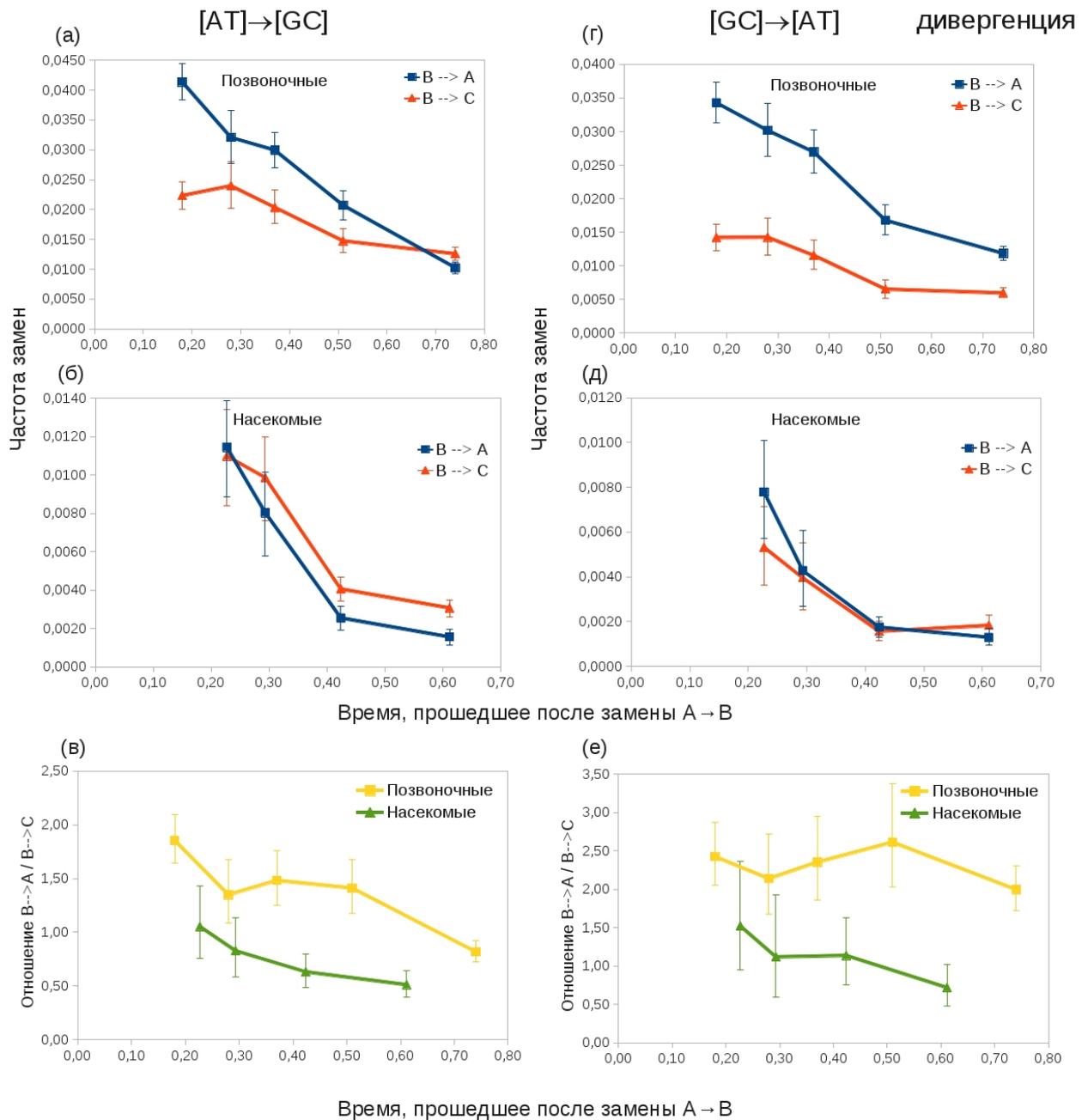


Рисунок 3.6. Зависимость частот замен от возраста предпочтительного в данный момент времени аллеля В, который зафиксировался ранее в результате замены А → В. а, б, в: замена АТ → GC во втором положении кодона; г, д, е: замена GC → АТ во втором положении кодона

## Полиморфизм

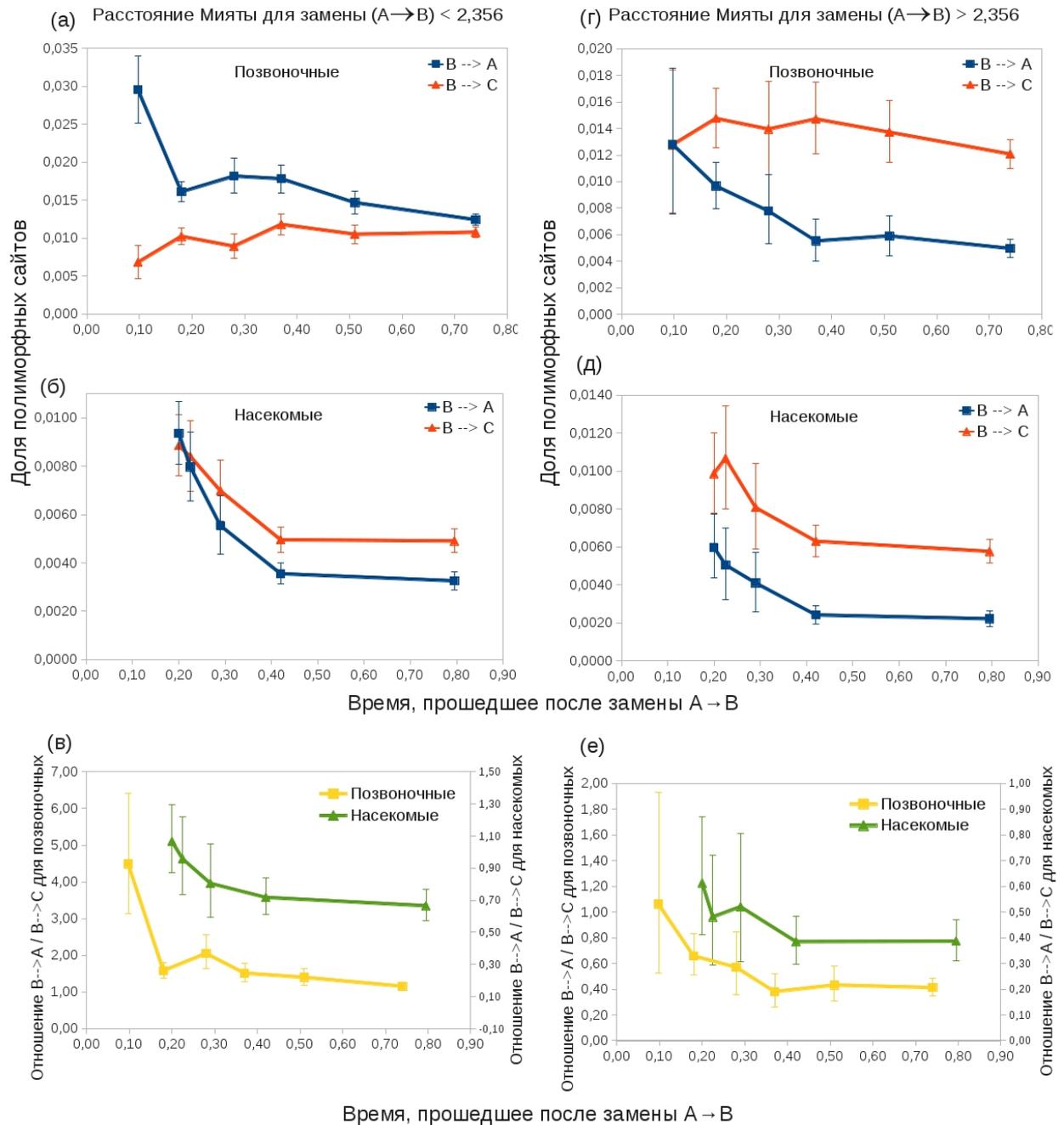


Рисунок 3.7. Зависимость уровня полиморфизма от возраста предпочтительного в данный момент времени аллеля В, который зафиксировался ранее в результате замены A→B. а,б,в: расстояние между аминокислотами (А,В) по шкале Мияты < 2,356; г,д,е: расстояние между аминокислотами (А,В) по шкале Мияты  $\geq 2,356$

## дивергенция

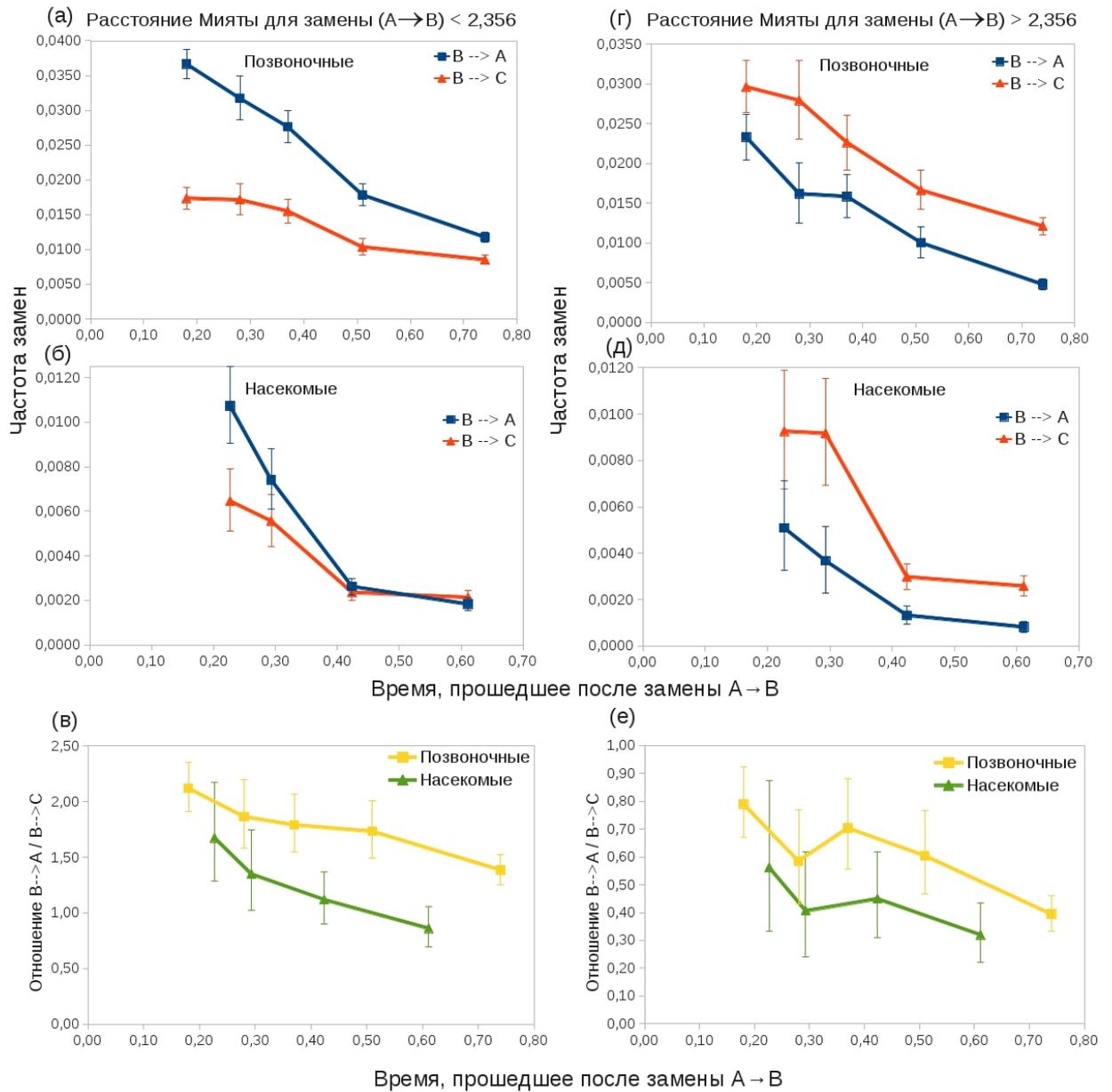


Рисунок 3.8. Зависимость частот замен от возраста предпочтительного в данный момент времени аллеля В, который зафиксировался ранее в результате замены A→B. а,б,в: расстояние между аминокислотами (А,В) по шкале Мияты < 2,356; г,д,е: расстояние между аминокислотами (А,В) по шкале Мияты ≥ 2,356

Сам по себе факт того, что частота полиморфизмов и замен аллеля В высока в тех сайтах, где В зафиксировался недавно, не обязательно свидетельствует об изменении адаптивного ландшафта. Предположим, что различным сайтам в геноме соответствуют различные ландшафты приспособленности, которые не меняются со временем. Тогда в набор сайтов, в которых наблюдается замена  $A \rightarrow B$ , с большей вероятностью будут входить сайты, в которых приспособленности А и В близки; поэтому в этих сайтах будет также наблюдаться избыток замен аллеля В. Падение отношений частот  $B \rightarrow A/B \rightarrow C$  совместимо с гипотезой о различных, но неизменных локальных ландшафтах приспособленности сайтов только в том случае, если репертуар аминокислот в быстро эволюционирующих сайтах будет узким (включающим аминокислоты А и В), а в медленно эволюционирующих сайтах репертуар должен быть широким. Однако верно противоположное (см. главу 4), поскольку быстро эволюционирующие сайты в основном не находятся под стабилизирующим отбором и могут принимать более широкий набор аминокислот.

Уменьшение частоты полиморфизмов  $B \rightarrow C$  и замен в зависимости от эволюционного времени, прошедшего после фиксации аллеля В, указывает на то, что приспособленность текущего аллеля В увеличивается со временем. Однако если бы это увеличение было единственным изменением ландшафта приспособленности после замены  $A \rightarrow B$ , а приспособленность А и С оставалась неизменной, отношение частот замен  $B \rightarrow A$  к  $B \rightarrow C$  и полиморфизмов должно было бы увеличиваться, а не уменьшаться со временем, в предположении, что в момент замены предковой аминокислоте А соответствовало большее значение приспособленности, чем аминокислотам С, которые никогда ранее не наблюдались (см. главу 2). Напротив, уменьшение приспособленности А приводит к наблюдаемой динамике, т.е. падению отношения  $B \rightarrow A/B \rightarrow C$ . Поэтому наблюдаемая динамика отношения  $B \rightarrow A/B \rightarrow C$  подтверждает, что, хотя оба эффекта — и увеличение приспособленности В, и уменьшение

приспособленности  $A$  – происходят после замены  $A \rightarrow B$ , второй эффект сильнее, и именно он должен быть основной причиной уменьшения скорости конвергентной эволюции с увеличением эволюционного расстояния [24, Рисунок 5] и уровня ревертирующих полиморфизмов и реверсий с возрастом прямой замены (Рисунок 3.4). Спустя время, соответствующее  $\sim 1$  нуклеотидной замене на синонимичный сайт после замены  $A \rightarrow B$ , приспособленность замененной аминокислоты  $A$  достигает нового, стабильно низкого значения.

### 3.3 Выводы

Изменение локального адаптивного ландшафта — медленный процесс, и, вероятно, он происходит в большой степени из-за замен в других сайтах белка или в других белках.

Поскольку в среднем отношение количества несинонимичных замен и замен в нефункциональных частях генома составляет  $\sim 0,1$ , белок “забывает” о замещенной аминокислоте после того, как  $\sim 20\%$  аминокислот в геноме заменились. Конечно, эти второстепенные замены не влияют напрямую на приспособленность замененной аминокислоты  $A$ . Их отрицательное влияние на приспособленность  $A$  должно происходить из-за преобладания отрицательного эпистаза: замена в произвольном сайте генома скорее снижает, чем увеличивает, приспособленность случайно выбранного генотипа с большой приспособленностью. Это не удивительно, поскольку большинство генотипов имеют низкие (нулевые) приспособленности. Для того, чтобы понять, являются ли эти замены во второстепенных сайтах по большей части адаптивными или околонеутральными, необходимы дополнительные исследования.

## 4 Оценка эффекта гетерогенности адаптивного ландшафта между сайтами

### 4.1 Эффект гетерогенности сайтов

При расчете частот эволюционных событий по файлам множественных выравниваний (глава 3) все сайты перечислены в линейном порядке, информация о пространственной структуре белка утрачена, следовательно, все сайты равноправно учитываются при подсчете числа событий. В действительности сайты не одинаковы по своим свойствам: в зависимости от положения сайта в пространственной структуре белка в нем может быть предпочтительна та или иная аминокислота.

Пусть каждый сайт белок-кодирующего гена принадлежит одному из двух классов с разными, но статическими, адаптивными ландшафтами (Рисунок 4.1). Ландшафт сайтов класса 1 более плоский, и в нем замена  $A \rightarrow V$  близка к нейтральной, а в ландшафте сайтов класса 2 замена  $A \rightarrow V$  более радикальна.

В силу статичности адаптивного ландшафта обратная замена  $V \rightarrow A$  будет более вероятной (будет происходить с большей скоростью) в сайтах класса 1, чем в сайтах класса 2. Следовательно, в среднем в сайтах класса 1 эволюционное расстояние между прямой и обратной заменой будет меньше, чем в сайтах класса 2.

Таким образом, наблюдаемое падение частоты обратных замен (полиморфизмов) с увеличением расстояния между прямой и обратной заменой (полиморфизмом) может быть следствием подразделения сайтов на классы при статическом адаптивном ландшафте, т.е. следствием зависимости адаптивного ландшафта от линейной пространственной координаты в выравнивании белок-кодирующего гена, а не следствием изменения адаптивного ландшафта со временем.

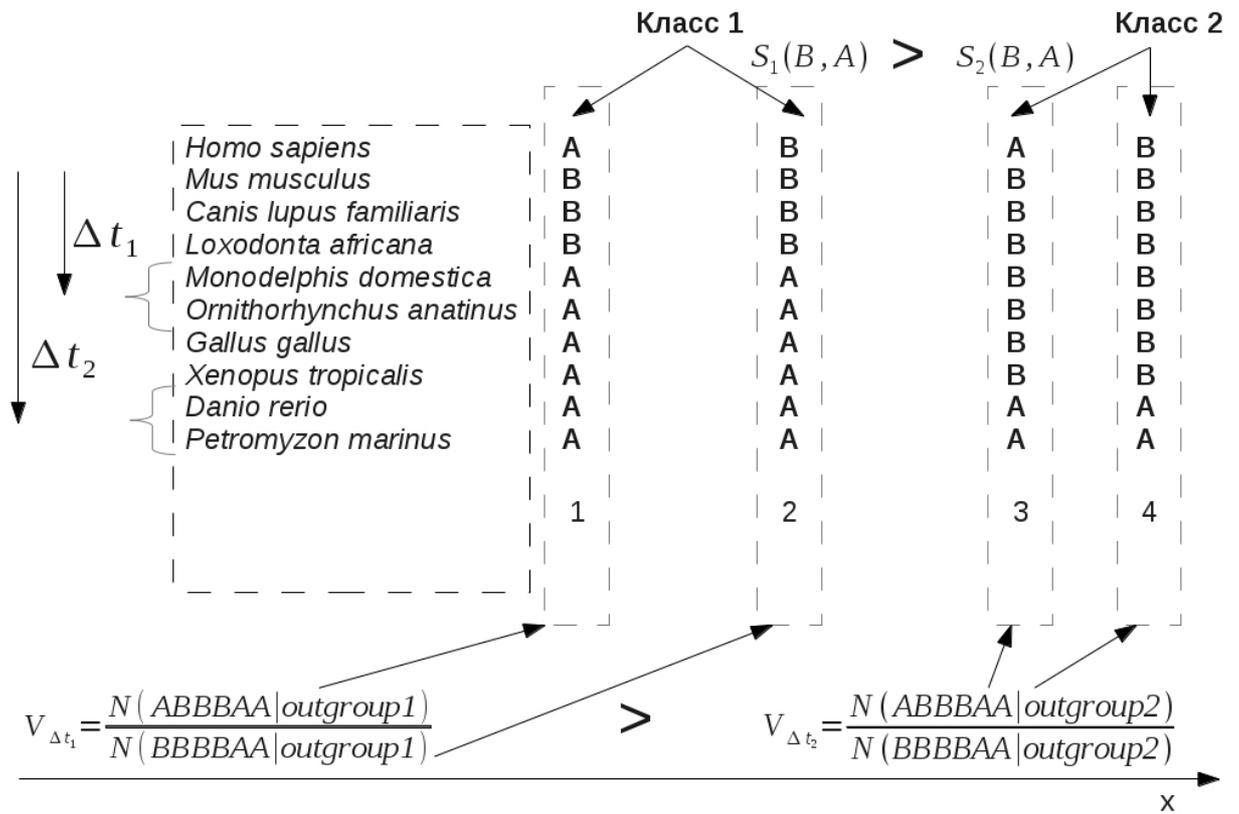


Рисунок 4.1. Классы сайтов в выравнивании белок-кодирующей последовательности. Показан пример наблюдаемых аминокислот для двух сайтов класса 1 и для двух сайтов класса 2.  $V$  — скорость реверсий  $A \rightarrow B \rightarrow A$ , равная отношению количества сайтов с реверсией к количеству сайтов, в которых произошла только прямая замена  $A \rightarrow B$ , для соответствующего эволюционного расстояния между прямой и обратной заменами ( $\Delta t$ ). Поскольку и прямая, и обратная замены в сайтах класса 1 соответствуют меньшим изменениям приспособленности, чем в сайтах класса 2, множество сайтов, в которых прямая замена наблюдалась недавно, будет обогащено сайтами класса 1, и  $V_{\Delta t_1} > V_{\Delta t_2}$ .

Назовем описанную гипотезу гипотезой гетерогенности сайтов. В действительности должны наблюдаться оба эффекта, влияющие на наблюдаемые паттерны — и изменчивость сайтов по форме адаптивного ландшафта, и зависимость ландшафта от времени. Вопрос в том, имеет ли

временной эффект значимую амплитуду на фоне эффекта разнообразия сайтов.

Для проверки гипотезы гетерогенности сайтов был использован ABC-тест. Исходим из того, что мы уже наблюдаем падение частоты замен  $B \rightarrow A$  с увеличением эволюционного расстояния между прямой заменой  $A \rightarrow B$  и обратной  $B \rightarrow A$  (глава 3). Будем наблюдать за отношением количества обратных замен из аминокислоты  $B$  в аминокислоту  $A$  к количеству боковых замен  $B \rightarrow C$ . Это отношение также убывает с увеличением эволюционного расстояния (Рисунок 3.4).

Пусть эффект гетерогенности сайтов преобладает; это значит, что большая частота замен  $B \rightarrow A$  и  $B \rightarrow C$  на меньших эволюционных расстояниях объясняется преимущественным попаданием сайтов в класс 1, в котором предпочтительнее обратные замены  $B \rightarrow A$ , а меньшая частота замен  $B \rightarrow A$  и  $B \rightarrow C$  объясняется преимущественным попаданием сайтов в класс 2, в котором предпочтительнее боковые замены  $B \rightarrow C$ . В таком случае в сайтах класса 2 должно наблюдаться повышенное количество боковых замен  $B \rightarrow C$  по сравнению с обратными заменами  $B \rightarrow A$ .

Количество боковых замен связано с количеством аминокислот, разрешенных в данном сайте — с шириной спектра аминокислот. Чем шире спектр возможных аминокислот, тем больше замен из аминокислоты  $B$  в аминокислоту, отличную от  $A$  и  $B$  (именно таково определение варианта  $C$ ). Следовательно, в сайтах класса 1 должен быть более узкий спектр разрешенных аминокислот; поэтому там чаще происходят обратные замены  $A \rightarrow B$ , по сравнению с сайтами класса 2.

В сайтах класса 1 наблюдается большая частота замен (скорость эволюции), чем в сайтах класса 2 (Рисунок 3.4в,е).

Таким образом, случае преобладания эффекта гетерогенности сайтов в сайтах с более узким спектром разрешенных аминокислот скорость эволюции

должна быть выше, чем в сайтах с более широким спектром. Однако ширина спектра должна быть положительно скоррелирована со скоростью молекулярной эволюции: чем больше разрешенных состояний в сайте, тем большее количество замен в нем должно происходить. Далее это предположение тестируется в эволюции белков плацентарных млекопитающих и насекомых.

#### 4.2 Две характеристики белок-кодирующего сайта

Эволюцию белкового сайта можно охарактеризовать, с одной стороны, скоростью (частотой) замен в этом сайте, а с другой стороны — шириной спектра допустимых в сайте аминокислот.

Предположим в духе нейтральной теории молекулярной эволюции [33], что в некотором сайте разрешены аминокислоты, которые дают практически одинаковый вклад в приспособленность, а другие аминокислоты запрещены, поскольку их вклад в приспособленность значительно ниже. Если количество разрешенных аминокислот велико, то для таких сайтов скорость эволюции будет высокой, а спектр эволюции — широким. Следовательно, можно ожидать, что скорость эволюции и ширина спектра наблюдаемых аминокислот по всем сайтам должны быть положительно скоррелированы. Задача настоящего раздела – исследовать эту корреляцию на основе имеющихся данных по полным геномам.

Пусть на филогении из  $N$  видов (не считая внешнего вида (аутгрупп), по которому определяется предковое состояние) (Рисунок 4.2.) в сайте произошло  $n$  независимых замен предковой аминокислоты на одну из  $k$  производных аминокислот. Замена каждого типа (замена на одну и ту же производную

аминокислоту) встречается  $x_i$  раз,  $\sum_{i=1}^k x_i = n$ , или с частотой  $p_i = \frac{x_i}{n}$ .

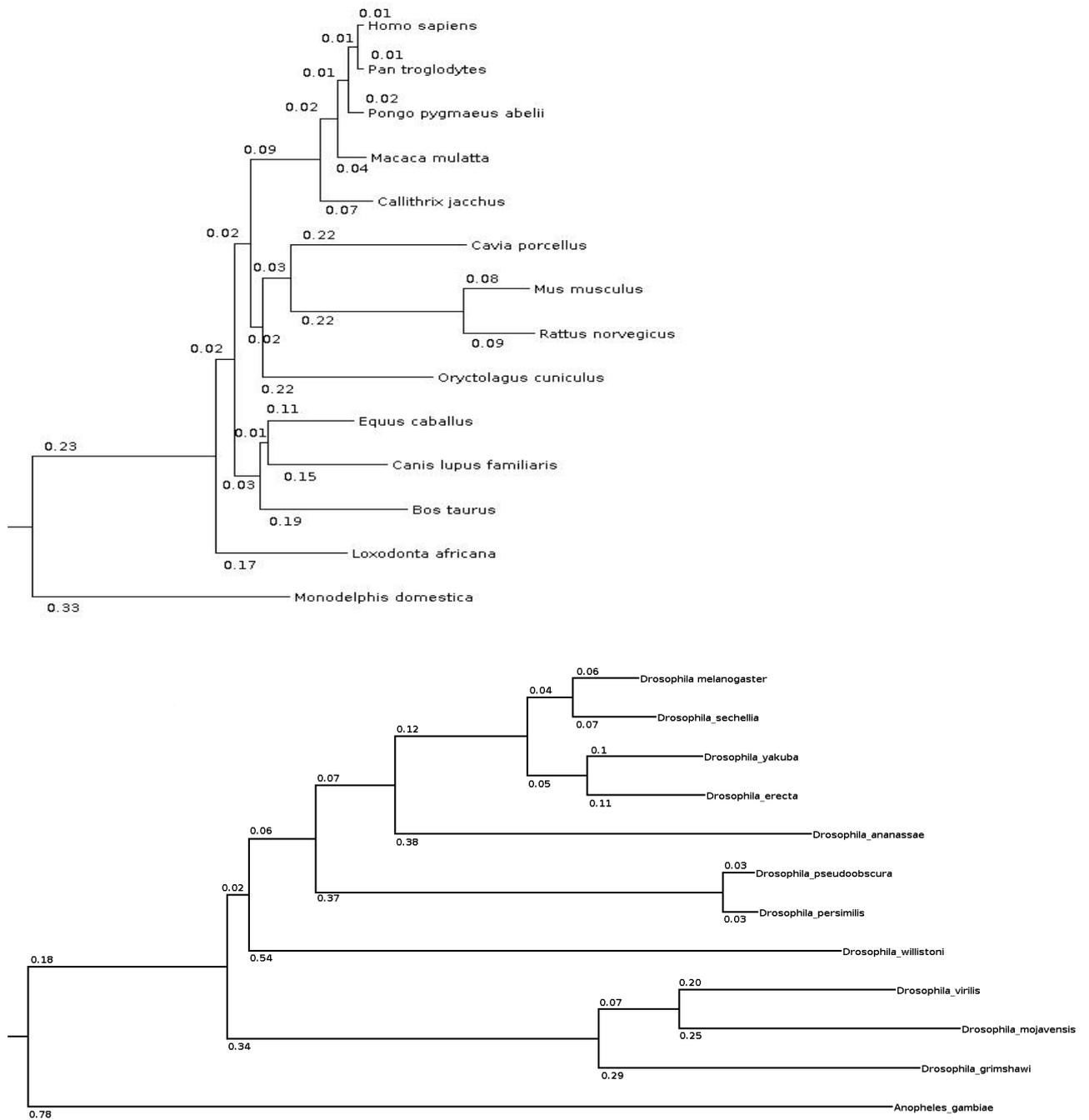


Рисунок 4.2. Филогенетические деревья плацентарных млекопитающих и дрозофил, использованные в анализе. Длины ветвей измерены в синонимичных заменах на синонимичный сайт ( $K_s$ )

### 4.2.1 Скорость эволюции

Скорость эволюции можно определять как количество аминокислотных замен, которые происходят за единицу астрономического времени. Можно также использовать тот факт, что скорость, с которой происходят селективно нейтральные замены, постоянна (при постоянности скорости мутирования [33]), и рассчитывать скорость эволюции как число замен на одну селективно нейтральную замену.

Охарактеризуем скорость эволюции в сайте величиной  $n$  – количеством независимых замен предковой аминокислоты. В случае  $N$  видов  $n$  лежит в промежутке от 0 до  $N$ .

### 4.2.2 Ширина спектра допустимых аминокислот

Ширину спектра эволюции — разнообразие аминокислот, которые могут занять сайт — можно определить как вероятность того, что две независимые замены предковой аминокислоты в сайте приводят к разным аминокислотам.

Ширину спектра допустимых аминокислот в сайте можно охарактеризовать величиной  $B$ , аналогичной несмещенной выборочной оценке ожидаемой гетерозиготности, которая широко используется в популяционной генетике [68]:

$$B = \frac{n}{n-1} \left( 1 - \sum_{i=1}^k p_i^2 \right).$$

Эта формула дает несмещенную оценку количества допустимых аминокислот в сайте.

$B$  можно также интерпретировать как долю пар замен предковой аминокислоты, таких, что производные аминокислоты отличны друг от друга.

Для  $n$  замен существует  $C_n^2 = \frac{n(n-1)}{2}$  пар (сочетаний по 2). Если среди  $n$  замен есть замены с одинаковым производным состоянием, то количество пар таких

замен составит  $\sum_{i=1}^k \frac{x_i(x_i-1)}{2}$  (количество сочетаний по 2 для каждого из множеств  $x_i$ ). Тогда долю пар замен с различными производными состояниями можно выразить как

$$\frac{\frac{n(n-1)}{2} - \sum_{i=1}^k \frac{x_i(x_i-1)}{2}}{\frac{n(n-1)}{2}} = \frac{\left( n^2 - n - \sum_{i=1}^k (x_i^2 - x_i) \right)}{n(n-1)} = \frac{n^2 - n - \sum_{i=1}^k x_i^2 + \sum_{i=1}^k x_i}{n(n-1)} = \frac{n^2 - \sum_{i=1}^k x_i^2}{n(n-1)}$$

Легко показать, что выражение (4.2) равно  $B$ . Преобразуем  $B$ :

$$B = \frac{n}{n-1} \left( 1 - \sum_{i=1}^k p_i^2 \right) = \frac{n \left( 1 - \sum_{i=1}^k \frac{x_i^2}{n^2} \right)}{n-1} = \frac{n^2 - \sum_{i=1}^k x_i^2}{n(n-1)}$$

$B$  может меняться в пределах от 0 (все наблюдаемые замены ведут к одной и той же аминокислоте, что означает, что лишь одна не-предковая аминокислота является допустимой) до 1 (все наблюдаемые замены являются заменами на разные аминокислоты, что означает максимальное разнообразие допустимых замен). Ясно, что  $B$  может быть определено только для сайтов с количеством замен не меньше двух. Мера разнообразия аминокислот, аналогичная  $B$ , была предложена в статье [69].

## 4.3 Методика расчета

### 4.3.1 Филогении и наборы данных

Отсутствие жёстко определённых требований к топологии филогении позволило нам использовать в этом разделе большее число видов, чем в

анализах из главы 3. Для анализа использовались два независимых набора данных: 13 полных геномов плацентарных млекопитающих (Рисунок 4.2) (все, секвенированные с покрытием больше 5x) с опоссумом *Monodelphis domestica* в качестве внешнего вида и 11 геномов видов плодовых мушек из рода *Drosophila* с малярийным комаром *Anopheles gambiae* в качестве внешнего вида.

Данные были загружены из базы данных UCSC Genome Browser [57]. Использовались выравнивания кодирующих последовательностей 45 видов позвоночных и 13 видов насекомых (<http://hgdownload.cse.ucsc.edu/goldenPath/hg19/multiz46way/alignments/knownCanonical.exonAA.fa.gz>). Для получения ортологичных наборов генов в соответствии с аннотацией *Homo sapiens* (*Drosophila melanogaster*) были извлечены белок-кодирующие гены. Нуклеотидные выравнивания ортологичных генов были транслированы в аминокислотные последовательности. Результирующие наборы данных содержали 3 705 042 сайтов (насекомые) и 12 520 688 (плацентарные млекопитающие).

#### 4.3.2 Реконструкция замен в предковых линиях

Для реконструкции предковых состояний и соответствующих замен был использован метод наибольшего правдоподобия, реализованный в программном пакете PAML [65]. По множественному выравниванию белок-кодирующего гена команда `codeml` дает список аминокислотных замен для каждого сайта. Схема расчетов приведена на Рисунке 4.3.

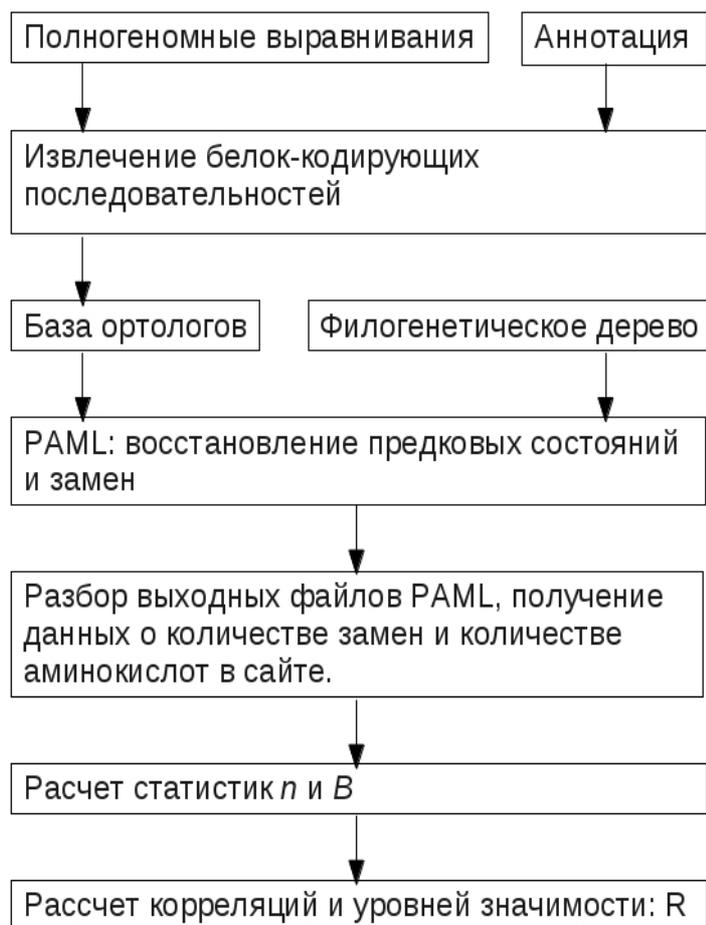


Рисунок 4.3. Схема проведения расчетов

Для статистического анализа и построения графиков использовался пакет R [70].

## 4.4 Полученные корреляции

### 4.4.1 Корреляции по сайтам

На Рисунке 4.4 показана связь между скоростью эволюции и шириной спектра для объединённой выборки всех сайтов из всех белков. Во многих сайтах не произошло замен, либо произошла одна замена. Количество таких сайтов приведено в Таблице 4.1. Для этих сайтов величина  $B$  не определена (они не показаны на Рисунке 4.4).

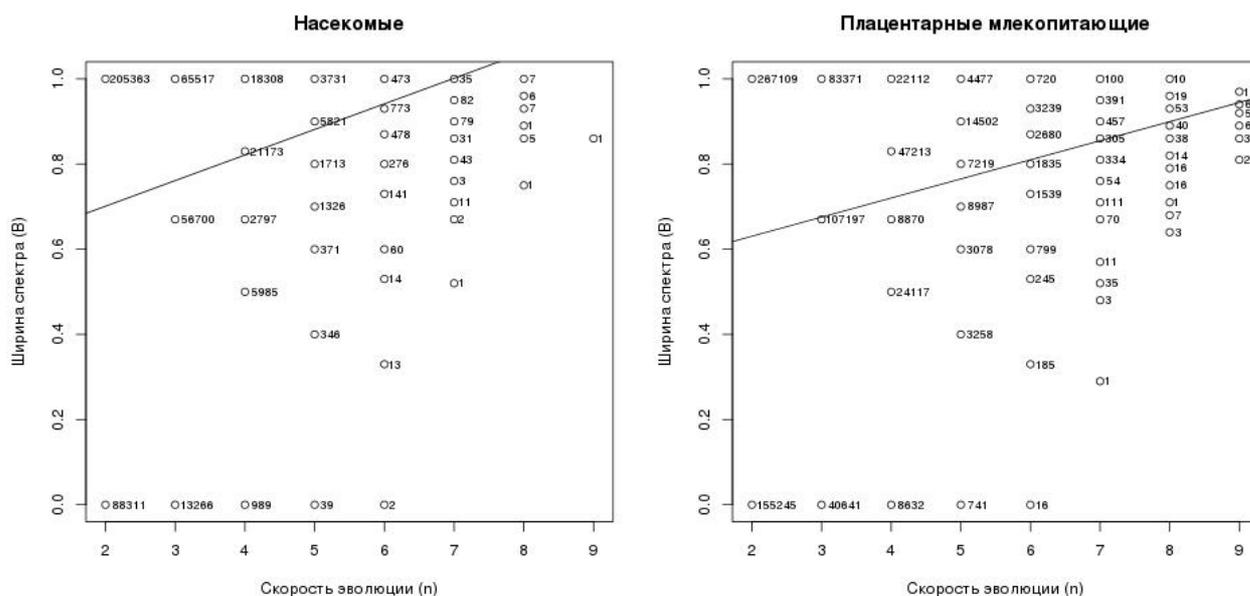


Рисунок 4.4. Количества сайтов для различных комбинаций значений  $n$  и  $B$  для *Drosophila* (слева) и для плацентарных млекопитающих (справа)

Таблица 4.1. Сайты с 0 и 1 заменой

	Количество сайтов (% от общего числа)	
	Инвариантных	С одной заменой
<i>Drosophila</i>	2 886 411 (78%)	324 330 (9%)
Плацентарные млекопитающие	5 086 661 (40%)	353 544 (3%)

#### 4.4.2 Корреляции с усреднением по белку

На Рисунке 4.5 показаны корреляции между  $n$  и  $B$  на уровне отдельных белков. Для каждого белка рассчитано арифметическое среднее величин, характеризующих скорость эволюции ( $n$ ) и ширину спектра ( $B$ ) для всех сайтов, в которых произошло как минимум две замены предковой аминокислоты.

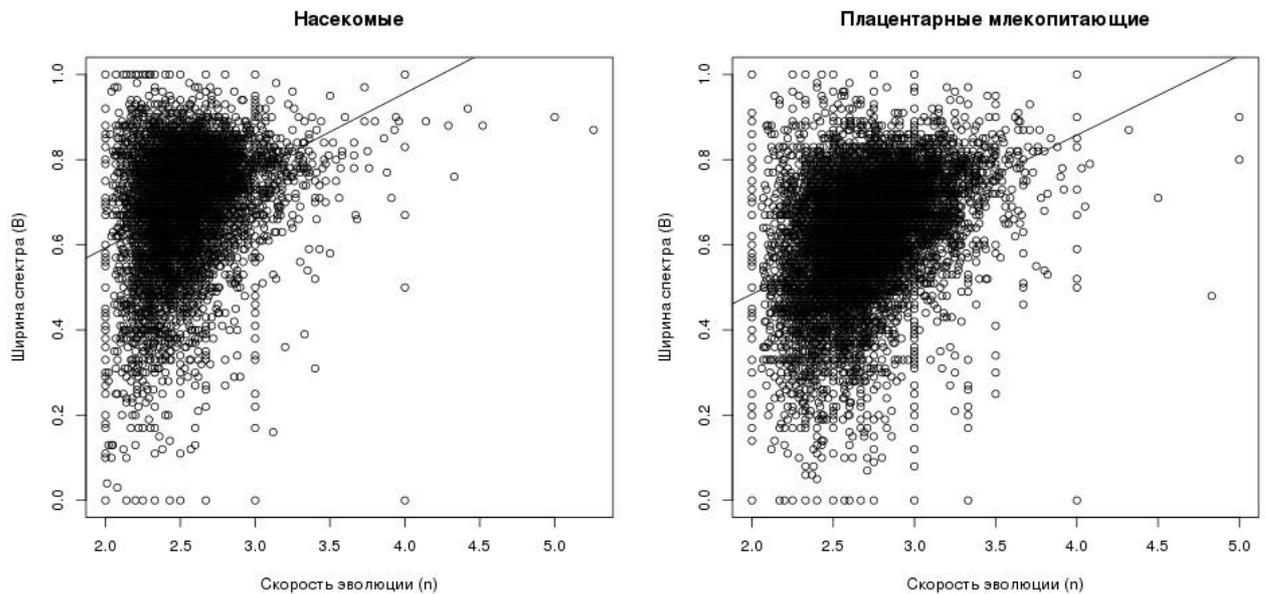


Рисунок 4.5. Скорость эволюции и ширина спектра, усредненные по белкам для *Drosophila* (слева) и плацентарных млекопитающих (справа)

В Таблице 4.2 приведены коэффициенты корреляции Пирсона и Спирмена между скоростью эволюции и шириной спектра, рассчитанные на основе полученных данных. Все коэффициенты значимо отличаются от 0 ( $p$ -value  $< 2,2 \cdot 10^{-16}$ ).

Таблица 4.2. Коэффициенты корреляции между скоростью эволюции и шириной эволюционного спектра для 20 аминокислот

	<i>Drosophila</i>		Плацентарные млекопитающие	
	R Пирсона	Rho Спирмена	R Пирсона	Rho Спирмена
По сайтам	0,12	-0,11	0,11	-0,12
По белкам	0,29	0,26	0,34	0,39

#### 4.4.3 Корреляции по классам аминокислот

На Рисунках 4.6, 4.7 и в Таблице 4.3 представлены аналогичные данные для случая, когда аминокислоты сгруппированы в 6 классов по физико-химическим свойствам в соответствии с работой [71] (AVLIMC, FWYH, STNQ, KR, DE, GP). Учитывались только те замены, в результате которых менялся класс аминокислоты в сайте. Все коэффициенты корреляции, представленные в Таблице 4.3, значимо отличаются от 0 ( $p\text{-value} < 2,2 \cdot 10^{-16}$ ).

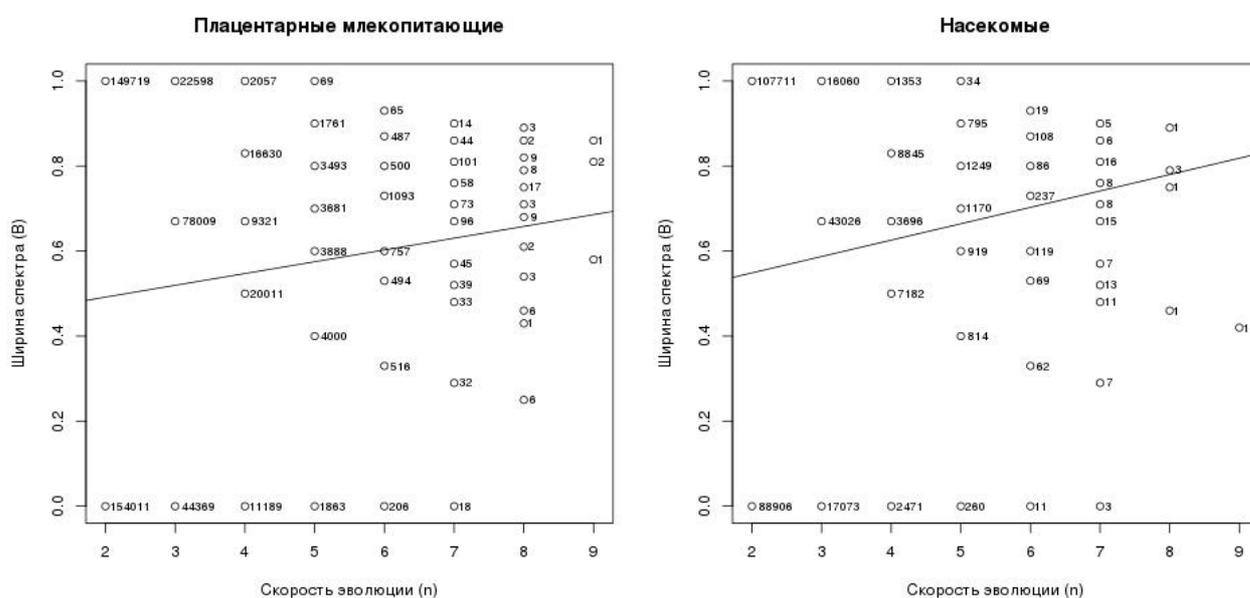


Рисунок 4.6. Количества сайтов для различных комбинаций значений  $n$  и  $B$  для *Drosophila* (слева) и для плацентарных млекопитающих (справа), рассчитанные с использованием классов сходства аминокислот

Таблица 4.3. Коэффициенты корреляции между скоростью эволюции и шириной эволюционного спектра для 6 классов аминокислот

	<i>Drosophila</i>		Плацентарные млекопитающие	
	R Пирсона	Rho Спирмена	R Пирсона	Rho Спирмена
По сайтам	0,06	-0,08	0,06	-0,05
По белкам	0,19	0,19	0,23	0,29

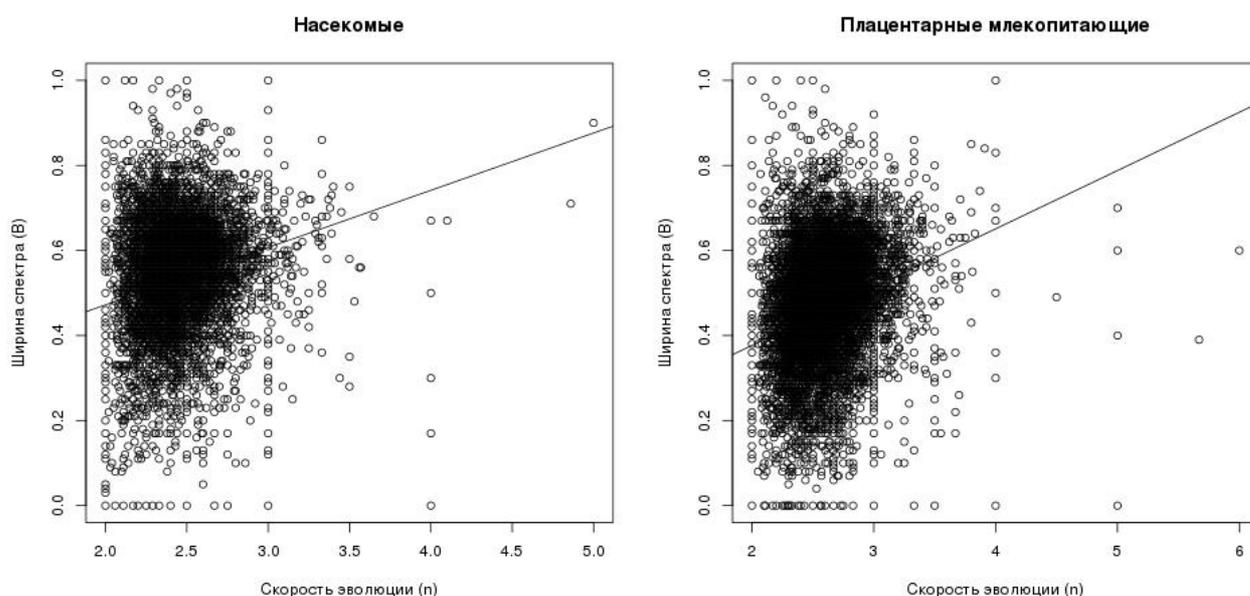


Рисунок 4.7. Скорость эволюции и ширина спектра, усредненные по белкам для *Drosophila* (слева) и плацентарных млекопитающих (справа), рассчитанные с использованием классов сходства аминокислот

#### 4.5 Выводы

Полученные данные показывают слабую положительную корреляцию между скоростью эволюции и шириной спектра как для индивидуальных сайтов, так и в среднем по белку, как для млекопитающих, так и для насекомых (Таблицы 4.2,4.3). Исключение составляют ранговые корреляции, рассчитанные по сайтам, которые очень слабо отрицательны. Корреляции усредненных по белку значений скорости эволюции и ширины спектра имеют большие значения, поскольку при усреднении уменьшается дисперсия значений ширины спектра и скорости эволюции.

Ранговые корреляции Спирмена для сайтов отрицательны при положительных корреляциях Пирсона. Это может объясняться структурой набора данных: для низких значений скорости эволюции (2-3 замены) присутствует множество точек со значением ширины спектра 0

(Рисунки 4.4, 4.6). Ранговая корреляция не учитывает разницу в значениях ширины спектра допустимых аминокислот, и при большом количестве сайтов с нулевым значением  $B$  может давать слабо отрицательные значения. Усреднение скоростей эволюции ширины спектра по белкам лишает набор данных этой структуры, и в этом случае значения коэффициентов корреляции Спирмена и Пирсона положительные.

В сайтах, которые эволюционируют быстрее, разнообразие допустимых аминокислот также немного выше. Простейшее объяснение этой положительной корреляции в том, что большинство аминокислотных замен происходит в селективно нейтральных, или практически нейтральных, сайтах [33]. Эти сайты эволюционируют со скоростью выше средней вследствие отсутствия селективного ограничения, и в них допустимы многие аминокислоты, что приводит к положительной корреляции между скоростью и шириной спектра.

Не наблюдается отрицательная корреляция скорости эволюции и ширины спектра, чего можно ожидать в предположении о преобладании эффекта гетерогенности сайтов, следовательно, изменение адаптивного ландшафта во времени имеет значимую амплитуду на фоне эффекта гетерогенности сайтов.

Однако полученная корреляция достаточно мала. В случае отдельных сайтов это может быть следствием малого количества видов в обоих наборах данных. В самом деле, для 11 или 13 генотипов значения скорости эволюции и ширины спектра определяются для каждого конкретного сайта только со значительной ошибкой, что снижает корреляцию. Однако эффект низкого размера выборки должен быть ослаблен, когда данные характеристики рассчитываются в среднем по белку, каждый из которых состоит из множества сайтов, и, действительно, корреляция на уровне белков значительно выше.

Суммарная длина всех веток филогенетического дерева в единицах  $Ks$  для плацентарных млекопитающих равна 3,20, а для дрозофил 2,08. Предполагая, что эволюция в синонимичных сайтах практически нейтральна, можно ожидать

~6 и ~4 несинонимичных замены в кодоне на этих деревьях в случае отсутствия селективного ограничения ( $K_s$  рассчитывается как количество синонимических замен на синонимический сайт, в случае отсутствия селективного ограничения замены в синонимических и несинонимических сайтах равновероятны, несинонимических сайтов примерно в 2 раза больше чем синонимических, т.к. замены в первом и втором положении кодона в основном несинонимические, а замены в третьем положении — в основном синонимические). Сайты с таким большим количеством замен редки (13% сайтов с 4 заменами и больше у дрозофил, 2% сайтов с 6 заменами и больше у млекопитающих) (Рисунок 4.4), что указывает на широко распространенное селективное ограничение на скорость эволюции. Значение ширины спектра  $B$  в сайте без ограничения должно быть близко к 1, поскольку для каждой аминокислоты в соответствии с генетическим кодом должно быть от 5 до 12 соседей, доступных в результате одной нуклеотидной мутации, при количестве замен в сайте равном 4-6 в среднем все замены будут различными. И снова мы видим более низкие значения, что указывает на значительное селективное ограничение на ширину спектра.

Таким образом, скорость эволюции и ширина спектра — это две различные и лишь слабо связанные характеристики белковой эволюции, которые нужно использовать отдельно в сложных анализах. Если доступно выравнивание большого количества ортологичных белков, так что значения для каждой из характеристик можно оценить с хорошей точностью для отдельных сайтов, было бы интересно изучить сайты, в которых скорость эволюции высока, а ширина спектра невелика, и наоборот. Сайты, которые эволюционируют очень быстро, но допускают лишь малое количество аминокислот, могут подвергаться эпизодам положительного отбора, а сайты, которые эволюционируют медленно, но допускают большое количество аминокислот, могут быть под эпистатическим отбором, на который влияет эволюция в других сайтах.

## **5 Размещение стоп-кодонов уменьшает количество нонсенс-мутаций**

Молекулярная эволюция белок-кодирующих последовательностей эукариот изучается на фоне достаточно консервативного аппарата трансляции, который был сформирован в эволюции ранней жизни.

При эволюционном формировании аппарата трансляции можно предположить одновременное существование организмов с различными генетическими кодами, обладающими неодинаковой эффективностью трансляции. Следовательно, можно представить себе адаптивный ландшафт над пространством генетических кодов, в котором каждому из возможных кодов присваивается некоторое значение приспособленности, соответствующее минимизации эффектов определенных мутаций.

В данной главе изучается такой адаптивный ландшафт и показывается положение на нем стандартного генетического кода.

### **5.1 Возникновение генетического кода**

Аминокислотная последовательность белка определяется нуклеотидной последовательностью кодирующего его гена в соответствии с генетическим кодом. Одну аминокислоту задает нуклеотидный триплет, или кодон — три последовательно идущих нуклеотида. Порядок нуклеотидов значим, поэтому существует  $4^3=64$  различных кодона. Из них в стандартном генетическом коде 61 является смысловым, кодирующим аминокислоты, а 3 — терминирующими, или стоп-кодонами, дающими сигнал к прекращению синтеза белка. С формальной точки зрения генетический код есть отображение алфавита из 64 триплетов с определенной внутренней структурой на множество, состоящее из 20 символов (аминокислот) и 1 знака препинания.

Реализовавшийся в природе стандартный генетический код

является одним из огромного числа возможных, однако он универсален, т.е. един для всех организмов (за несколькими исключениями, затрагивающими, как правило, относительно небольшое число видов).

Проблема эволюционного происхождения генетического кода, и более широко, всего аппарата трансляции важна для понимания хода ранней эволюции биологической жизни. Исчерпывающего описания этого процесса, по всей видимости, получить не удастся, так как фиксация кода произошла на ранних стадиях эволюции, и на последующих стадиях происходили изменения геномов и механизмов регуляции; однако многие аспекты удалось выяснить с высокой степенью уверенности [72, 73].

Сразу после расшифровки стандартного генетического кода появилась теория “замороженного кода” Ф. Крика (F. Crick) [74], согласно которой зафиксировался код, не обладающий адаптивными свойствами. Уже само по себе наличие кода, даже и неоптимального, является эволюционным преимуществом.

Каждое изменение в коде влечёт за собой множественные замены в белках, поэтому сама возможность эволюции кода некоторое время ставилась под сомнение. Открытие митохондриального кода и других девиантных кодов доказало, что соответствие кодонов аминокислотам может меняться. Девиантные генетические коды, по всей видимости, являются модификациями стандартного [75].

Одной из первых теорий возникновения кода стала стереохимическая теория, согласно которой имеется пространственное соответствие между молекулами, представляющими соответствующие триплеты, и аминокислотами [76].

Было обнаружено отклонение распределения аминокислот по генетическому коду от случайного: близкие по гидрофобности аминокислоты кодируются близкими кодонами [77]. Однако стандартный код не оптимален по этому критерию, а субоптимален, и лучшие коды легко находятся простейшим

эволюционным алгоритмом [78].

Большинство аминокислот не синтезируются абиотически, и аминокислоты, входящие в одни и те же биохимические пути, имеют тенденцию кодироваться триплетами, которые начинаются с одинакового нуклеотида. На этом основана коэволюционная теория возникновения генетического кода: первичный код был еще более вырожденным, и постепенно, параллельно с развитием биосинтеза аминокислот, добавлялись новые аминокислоты. Построены теоретические модели такого процесса [79], которые позволяют получить вырожденный и устойчивый к ошибкам код за счет коэволюции.

По-видимому, различные сценарии происхождения кода не противоречат друг другу, а могли действовать либо сменяя друг друга последовательно, либо одновременно, причем степень влияния каждого из сценариев могла меняться во времени [80].

## 5.2 Разметка генетического кода

Назовем разметкой генетического кода способ деления кодонов на смысловые и терминирующие. Разметку стандартного кода будем называть стандартной разметкой. Для нее терминирующими являются триплеты TAA, TAG и TGA. Для разметок с тремя стоп-кодонами существует  $C_{64}^3 = 41\ 664$  варианта.

Признаки начала и конца нуклеотидного триплета в мРНК отсутствуют. Трансляция происходит последовательно, триплет за триплетом. С одной и той же мРНК можно считать три различные белковые последовательности: в рамках 0, +1, -1 (в зависимости от положения первого нуклеотида). В подавляющем большинстве случаев функциональный белок синтезируется только в одной рамке считывания. Трансляция со сдвигом рамки часто приводит к преждевременной терминации белковой цепи.

Пару смысловых кодонов, которая при трансляции со сдвигом дает терминирующий кодон, будем называть запретной. Такова, например, пара АТА–GCA, поскольку она даёт стоп-кодон TAG при чтении в рамке +1.

Смысловые кодоны, соседствующие с терминирующими и вследствие этого подверженные точечным нонсенс-мутациям, будем называть уязвимыми кодонами.

### 5.3 Критерии адаптации для расположения стоп-кодонов

Разметка генетического кода влияет на вероятность возникновения стоп-кодонов при чтении со сдвигом рамки (при условии, что триплетный состав последовательности однороден) и на количество кодонов, подверженных точечной нонсенс-мутации, а также на общее количество таких мутаций.

Критерий 1. Сдвиг рамки считывания — это нелокальная ошибка, полностью меняющая аминокислотную последовательность белка. По всей видимости, преимущество будет иметь такой аппарат трансляции, который возможно раньше прерывает чтение за счет появления стоп-кодона в новой рамке. Следовательно, оптимальной является разметка, максимизирующая вероятность появления стоп-кодона, или, иными словами, разметка с максимальным количеством запретных пар кодонов.

Критерий 2. Нонсенс-мутация приводит к преждевременному прекращению трансляции. Частично сформированная последовательность белка может привести к ошибкам образования вторичной, третичной структур белка и потере функциональности. Согласно этому критерию оптимальна разметка, минимизирующая вероятность нонсенс-мутаций.

Критерий 2 распадается на 2 случая. Возможна минимизация суммарного (по всему коду) числа нонсенс-мутаций, либо минимизация количества уязвимых кодонов.

## 5.4 Положение разметки универсального генетического кода на ландшафте всех кодов

### 5.4.1 Устойчивость к мутациям сдвига рамки

Существует  $61 \times 61 = 3721$  пара смысловых кодонов — последовательностей из шести букв из набора {A;G;C;T}, ни одна из трехбуквенных половинок которых, стоящих в положениях 1–2–3 и 4–5–6, стоп-кодоном не является. Пара смысловых кодонов запретна, если она содержит три подряд идущих буквы, образующих стоп-кодон, в положениях 2–3–4 или 3–4–5.

Максимально возможное число запретных пар для каждого типа сдвига рамки (+1, -1) равно  $192 = 3 \times 4^3$  (3 варианта стоп-кодона  $\times$   $4^3$  варианта остальных нуклеотидов). Причем максимум может достигаться для обоих типов сдвига только одновременно. В самом деле, необходимым и достаточным условием оптимальности разметки согласно критерию 1 является отсутствие таких стоп-кодонов, что конец одного из них является продолжением другого (например, TAG и GCT или TAG и AGC) или самого себя (например, CCC или GTG).

Легко убедиться, что стандартная разметка (со стоп-кодонами TAA, TAG, TGA) удовлетворяет этому условию, т.е. является оптимальной, давая 192 запретные пары кодонов в случае обоих сдвигов рамки. Полный перебор вариантов показывает, что всего существует 2 432 (5,8%) таких разметок. Распределение разметок по количеству запретных пар кодонов для сдвига -1 показано на Рисунке 5.1 Распределение для сдвига +1 выглядит точно так же, причем ни для одной разметки число запретных пар для сдвигов -1 и +1 не отличается более чем на 3.

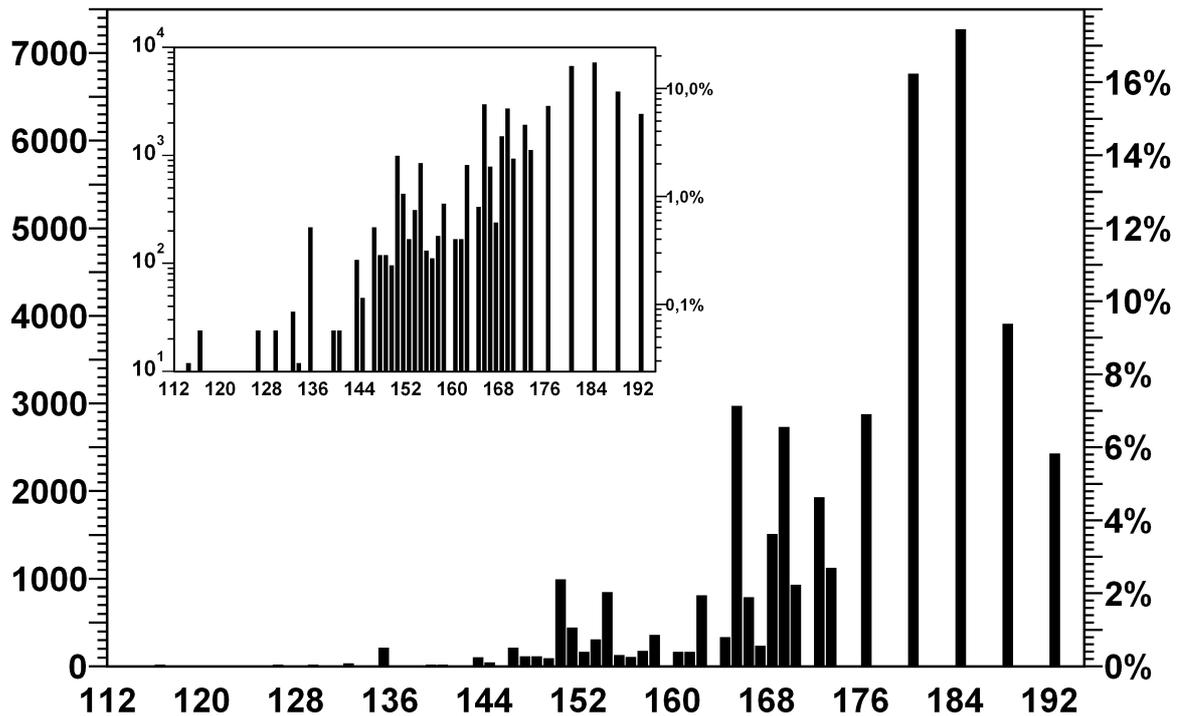


Рисунок 5.1. Распределение разметок по числу запретных пар кодонов. По оси абсцисс отложено количество запретных пар кодонов, по осям ординат — количество разметок с данным значением количества запретных пар и их доля. На врезке — те же данные с логарифмическим представлением по оси ординат

#### 5.4.2 Устойчивость к нонсенс-мутациям

С точки зрения устойчивости к точечным мутациям разметки можно разбить на 4 группы, определяемые взаимным соседством стоп-кодонов:

В группе 1 каждый стоп-кодон является соседним (т.е. получаемым заменой одного нуклеотида) по отношению к двум другим. В этом случае 1 смысловой кодон является соседом всех трех стоп-кодонов и еще по 6 смысловых кодонов соседствуют с каждым из них (2 положения в триплете, по которым совпадают стоп-кодоны,  $\times$  3 основания, отличных от стоящих в этих положениях). Итого получается  $1+6+6+6 = 19$  уязвимых кодонов и  $1 \times 3 + 6 + 6 + 6 = 21$  возможная нонсенс-мутация. Последнее значение является минимально возможным.

Всего существует 27 мутаций, превращающих некие кодоны в

терминирующие (3 стоп-кодона x 3 положения в триплете x 3 основания, на которые заменяется основание в мутирующем кодоне). Каждая точечная мутация, превращающая один стоп-кодон в другой, уменьшает число мутаций, идущих на то, чтобы превращать в них смысловые кодоны. Количество мутаций, переводящих стоп-кодоны друг в друга, не может быть более 6; в данном случае их именно столько.

В группе 2 один стоп-кодон соседствует с двумя другими, но те, в свою очередь, друг другу соседями не являются. Именно такова стандартная разметка (стоп-кодоны: TAA, TAG и TGA — первый соседствует с остальными). Мутаций, переводящих стоп-кодоны друг в друга, здесь уже только 4, соответственно, на смысловую часть кода остаются 23 нонсенс-мутации. В этом случае найдется 5 смысловых кодонов, соседствующих сразу с двумя стоп-кодонами (TAC, TAT, TCA, TTA, TGG), и еще  $23 - 5 \times 2 = 13$ , которые являются соседями только одного стоп-кодона (AAA, AAG, AGA, GAA, GAG, GGA, CAA, CAG, CGA, TCG, TTA, TCG, TTA). Итого:  $5 + 13 = 18$  уязвимый кодон, что, как показывает анализ всех вариантов, является минимально возможным значением. Всего эта группа содержит 1 728 разметок, или 4,15% от общего числа.

В группе 3 два стоп-кодона являются соседями, а третий не соседствует ни с одним из них. Мутаций, переводящих стоп-кодоны друг в друга, остается всего 2, что дает 25 нонсенс-мутаций. Число уязвимых кодонов варьируется от 20 до 23.

В группе 4 соседних кодонов среди стоп-кодонов нет. Все 27 нонсенс-мутаций приходятся на смысловую часть кода. Число уязвимых кодонов — от 21 до 27.

## 5.5 Выводы

Стандартная разметка генетического кода обеспечивает максимальную вероятность появления стоп-кодона при чтении со сдвигом рамки на однородной по кодонному составу последовательности, а также минимальное количество кодонов, подверженных точечным нонсенс-мутациям. Всего существует 528 (1,3%) разметок с аналогичными свойствами.

Оптимальность стандартного генетического кода предполагает эволюционный сценарий его происхождения. Вероятно, в критерии приспособленности ранних организмов входила устойчивость аппарата трансляции к ошибкам. Фиксация стоп-кодонов, минимизирующая ошибки, связанные со сдвигом рамки считывания, могла повысить устойчивость раннего аппарата трансляции.

Стандартная разметка обеспечивает минимальное количество уязвимых для точечных нонсенс-мутаций кодонов, а не общее количество нонсенс-мутаций по коду. Критерием приспособленности раннего аппарата трансляции могло быть количество кодонов, расположенных на расстоянии более одной мутации от стоп-кодонов. Достижение генетического кода, оптимального по этому критерию, могло способствовать дальнейшей специализации неязвимых кодонов.

## Основные результаты и выводы

1. В молекулярной эволюции белок-кодирующих последовательностей позвоночных и насекомых после аминокислотной замены происходит изменение адаптивного ландшафта аминокислотного сайта: приспособленность предкового аллеля уменьшается со временем до фонового значения.
2. Вследствие изменения адаптивного ландшафта со временем, прошедшим после аминокислотной замены, наблюдается уменьшение частот обратных полиморфизмов и замен. Эта динамика может быть также обусловлена изменением адаптивного ландшафта во времени, состоящем в увеличении приспособленности производного аллеля, или неравномерностью адаптивного ландшафта по соотношению приспособленностей аллелей в разных аминокислотных сайтах (эффект гетерогенности сайтов).
3. Увеличение приспособленности производного аллеля должно приводить к увеличению отношения частоты обратных замен к частоте замен на аминокислоты, отличные от предковой и производной (боковых замен) со временем, прошедшим после прямой замены.
4. Уменьшение приспособленности предкового аллеля должно приводить к уменьшению отношения частоты обратных замен к частоте боковых замен со временем, прошедшим после прямой замены.
5. В эволюции позвоночных и насекомых наблюдается уменьшение отношения частоты прямых замен к частоте боковых замен, из чего следует преобладание эффекта уменьшения приспособленности предкового аллеля по сравнению с эффектом увеличения приспособленности производного аллеля.
6. В эволюции позвоночных и насекомых наблюдается слабая

положительная корреляция ширины спектра допустимых аминокислот и скорости молекулярной эволюции, из чего следует преобладание эффекта временной динамики адаптивного ландшафта над эффектом гетерогенности сайтов.

7. Стоп-кодны в генетическом коде расположены таким образом, что обеспечивают устойчивость к мутациям сдвига рамки считывания и нонсенс-мутациям.

## Благодарности

Автор выражает благодарность Г.А. Базыкину и А.С. Кондрашову за совместную работу и обсуждение вопросов динамики адаптивного ландшафта, Г.Г. Малинецкому, А.В. Подлазову и М.С. Бурцеву — за совместную работу и обсуждение задачи о разметке генетического кода, коллегам Р.Н. Ариффулову, О.А. Вахрушевой, М.А. Барановой, М.- Г.М. Зульпукарову, П.В. Куракину, Е.В. Лёушкину, С.А. Махову, Н.А. Митину, А.Д. Неверову, Н.В. Поповой, В.Б. Сеплярскому, Н.В. Терехановой, Д.В. Серебрякову — за плодотворные обсуждения, М.С. Гельфанду и А.А. Миронову — за замечательную исследовательскую атмосферу, созданную в ИППИ РАН и на ФББ МГУ, своей семье и друзьям, а особенно жене Ю.П. Бибило — за поддержку.

## Список литературы

1. Hamilton, M.B. Population genetics / M.B. Hamilton. – Wiley-blackwell, 2009.
2. Crow, J.F. An introduction to population genetics theory / J.F. Crow, M. Kimura. – Reprint of 1970 edition. – Harper and Row, 2009.
3. Hartl, D.L. Principles of population genetics / D.L. Hartl, A.G. Clark. – 3rd ed. – Sunderland: Sinauer Associates, 1997.
4. Bergstrom, C.T. Evolution / C.T. Bergstrom, L.A. Dugatkin. – W.W. Norton & Company, 2012.
5. Ridley, M. Evolution / M.Ridley. – 3rd ed. – Blackwell publishing, 2004.
6. Futuyma, D.J. Evolution / D.J. Futuyma. – Sunderland : Sinauer, 2005.
7. Kondrashov, A.S. Evolution of life (unpublished).
8. Orr, H.A. Genetic theories of adaptation: a brief history / H.A. Orr // Nature Reviews. Genetics. – 2005. – V 6. – P.119–127.
9. Gavrilets, S. Fitness landscapes and the origin of species / S. Gavrilets. – Princeton University Press, 2004.
10. Orr, H.A. Fitness and its role in evolutionary genetics / H.A. Orr // Nature Reviews Genetics. – 2009. – V.10. – P.531–539.
11. The adaptive landscape in evolutionary biology / Edited by E. Svensson and R. Calsbeek. – Oxford University Press, 2012.
12. Wright, S. Evolution in Mendelian Populations / S. Wright // Genetics. – 1931. – V. 16. – P.97–159.
13. Wright, S. The roles of mutation, inbreeding, crossbreeding, and selection in evolution / S. Wright // Proceedings of the Sixth International Congress on Genetics. – 1932. – P.355–366.
14. Wright, S. Surfaces of selective value revisited / S. Wright // The American Naturalist. – 1988. – V.131. – N 1. – P.115–123.

15. Philipps, P.C. Epistasis – the essential role of gene interactions in the structure and evolution of genetic systems / P.C. Philipps // *Nature Reviews. Genetics.* – 2008. – V.9. – P.855–867.
16. Lewontin, R.C. The evolutionary dynamics of complex polymorphisms / R.C. Lewontin and K. Kojima // *Evolution.* – 1960. – V.14. – N 4. – P.458– 472.
17. Lewontin, R.C. Interaction between inversion polymorphisms of two chromosome pairs in the grasshopper *Moraba scurra* / R.C. Lewontin and M.J.D. White // *Evolution.* – 1960. – V.14. – N 1 – P.116–129.
18. Левонтин, Р. Генетические основы эволюции / Р. Левонтин. – Пер. с англ. – М.: Мир, 1978. – 351с., ил.
19. Fisher, R.A. The genetical theory of natural selection / R.A. Fisher. – Oxford, 1930.
20. Price, G.R. Fisher's 'fundamental theorem' made clear / G.R. Price // *Ann. Hum. Genet. Lond.* – 1972. – V. 36. – P.129–140.
21. Frank, S.A. Fisher's fundamental theorem of natural selection / S.A. Frank and M. Slatkin // *Trends in Evology and Evolution.* – 1992. – V.7. – N 3. – P.92– 95.
22. Frank, S.A. Wright's adaptive landscape versus Fisher's fundamental theorem / S.A. Frank // *The adaptive landscape in evolutionary biology* / Edited by Svensson E. and Calsbeek R. – Oxford University Press, 2012.
23. Smith, J.M. Natural selection and the concept of a protein space / J.M. Smith // *Nature.* – 1970. – V. 225. – P.563–564.
24. Povolotskaya, I.S. Sequence space and the ongoing expansion of the protein universe / I.S. Povolotskaya and F.A. Kondrashov // *Nature.* – 2010. – V.465. – P.922–926.
25. Breen, M.S. Epistasis as the primary factor in molecular evolution / M.S. Breen, C. Kemena, P.K.Vlasov, C. Notredame, and F.A. Kondrashov // *Nature.* – 2012. – V.490.– P.535–538.

26. Gillespie, J.H. Molecular evolution over the mutational landscape / J.H. Gillespie // *Evolution*. – 1984. – V.38. – N 5. – P.1116-1129.
27. Gillespie, J.H. The causes of molecular evolution / J.H. Gillespie. – Oxford University Press, 1994.
28. Wright, S. Random drift and the shifting balance theory of evolution / S. Wright // *Mathematical topics in population genetics* / Edited by K. Kojima. – Berlin: Springer-Verlag, 1970.
29. Bazykin, G.A. Positive selection at sites of multiple amino acid replacements since rat-mouse divergence / G.A. Bazykin, F.A. Kondrashov, A.Y. Ogurtsov, S. Sunyaev, and A.S. Kondrashov // *Nature*. – 2004. – V.429. – P.558–562.
30. Bazykin, G.A. Bursts of nonsynonymous substitutions in HIV-1 evolution reveal instances of positive selection at conservative protein sites / G.A. Bazykin, J. Dushoff, S.A. Levin, and A.S. Kondrashov // *Proc. Natl. Acad. Sci. USA*. – 2006. – V.103. – N 51. – P.19396–19401.
31. Kimura, M. Evolutionary rate at the molecular level / M. Kimura // *Nature*. – 1968. – V.217. – P.624–626.
32. King, J.L. Non-Darwinian evolution / King J.L and Jukes T.H. // *Science*. – 1969. – V.164. – P.788–798.
33. Кимура, М. Молекулярная эволюция: теория нейтральности / М. Кимура. – Пер. с англ. – М.: Мир, 1985. – 394с., ил.
34. Ohta, T. The nearly neutral theory of molecular evolution / T. Ohta // *Annu. Rev. Ecol. Syst.* – 1992. – V.23. – P.263–286.
35. Ohta, T. Development of neutral and nearly neutral theories / T. Ohta and G.H. Gillespie // *Theoretical population biology*. – 1996. – V.49. – P.128–142.
36. Akashi, H. Weak selection and protein evolution / H. Akashi, N. Osada, and T. Ohta // *Genetics*. – 2012. – V.192. – P.15–31.
37. Kauffman, S. Towards a general theory of adaptive walks on rugged landscapes / S. Kauffman and S. Levin // *J.Theor.Biol.* – 1987. – V.128. – P.11–45.

38. Kauffman, S.A. The origins of order. Self-organization and selection in evolution / S.A. Kauffman. – New York : Oxford University Press, 1993.
39. Kauffman, S.A. The NK model of rugged fitness landscapes and its application to maturation of the immune response / S.A. Kauffman and E.D. Weinberger // J. Theor. Biol. – 1989. – V.141. – P.211–245.
40. Bell, G. Fluctuating selection: the perpetual renewal of adaptation in variable environments / G. Bell // Phil. Trans. R. Soc. B. – 2010. – V.365. – P.87–97.
41. Kimura, M. Process leading to quasi-fixation of genes in natural populations due to random fluctuation of selection intensities / M. Kimura // Genetics. – 1954. – V.39. – P.280–295.
42. Gillespie, J.H. Natural selection with varying selection coefficients – a haploid model / J.H. Gillespie // Genetical Research. – 1973. – V.21. – N 2. – P.115– 120.
43. Mustonen, V. Adaptations to fluctuating selection in *Drosophila* / V. Mustonen and M. Lassig // Proc. Natl. Acad. Sci. USA. – 2007. – V.104. – N 7. – P.2277 –2282.
44. Sella, G. The application of statistical physics to evolutionary biology / G. Sella, A.E. Hirsh // Proc. Natl. Acad. Sci. USA. – 1995. – V. 102. – N 27. – P.9541– 9546.
45. Mustonen, V. From fitness landscapes to seascape: non-equilibrium dynamics of selection and adaptation / V. Mustonen, M. Lassig // Trends in Genetics. – 2009. – V. 25. – N 3. – P.111–119.
46. Mustonen, V. Fitness flux and ubiquity of adaptive evolution / V. Mustonen and M. Lassig // Proc. Natl. Acad. Sci. USA. – 2010. – V.107. – N 9. – P.4248 – 4253.
47. Gavrillets, S. Evolution and speciation on holey adaptive landscapes / S. Gavrillets // Trends in Ecology & Evolution. – 1997. – V. 12. – P.307–312.
48. Holland, J.H. Adaptation in natural and artificial systems / J.H. Holland. – University of Michigan Press, 1975.

49. Mitchell, M. An introduction to genetic algorithms / M. Mitchell. – The MIT Press, 1999.
50. Engelbrecht, A.P. Computational intelligence. An introduction. / A.P. Engelbrecht. – 2nd ed. – Wiley, 2007.
51. Bioinformatics for high throughput sequencing. / Edited by N. Rodriguez-Ezpeleta, M. Hackenberg, A.M. Aransay. – Springer, 2012.
52. Benson, D.A. Genbank / D.A. Benson, M. Cavanaugh, K. Clark, I. Karsch-Mizrachi, D.J. Lipman, J. Ostell, and E.W. Sayers // Nucleic Acids Res. – 2013. – V.41(D1). – D36–D42.
53. Growth of GenBank and WGS. – Режим доступа: <http://www.ncbi.nlm.nih.gov/genbank/statistics>.
54. Blount, Z. D. Genomic analysis of a key innovation in an experimental *Escherichia coli* population / Z.D. Blount, J.E. Barrick, C.J. Davidson, and R.E. Lenski // Nature. – 2012. – V. 489. – P.513–518.
55. Fasan, R. Evolutionary history of a specialized p450 propane monooxygenase / R. Fasan, Y.T. Meharena, C.D. Snow, T.L. Poulos, F.H. Arnold // J. Mol. Biol. – 2008. – V.383. – N 5. – P.1069–1080.
56. Extremophiles handbook / Edited by K. Horikoshi. – Springer, 2011.
57. UCSC genome bioinformatics. – Режим доступа: <http://genome.ucsc.edu/>.
58. The 1000 genomes project consortium. A map of human genome variation from population-scale sequencing. // Nature. – 2010. – V.467. – P.1061–1073.
59. The drosophila melanogaster genetic reference panel. // Nature. – 2012. – V.482. – P.173–178.
60. The Cancer Genome Atlas. – Режим доступа: <http://cancergenome.nih.gov/>.
61. McDonald, J.H., Kreitman, M. Adaptive protein evolution at the Adh locus in Drosophila / J.H. McDonald, M. Kreitman // Nature. – 1991. – V.351. – P.652– 654.
62. Bazykin, G.A. Detecting past positive selection through ongoing negative selection. / G.A. Bazykin and A.S. Kondrashov // Genome Biol. Evol. –

2011. – V.3. – P.1006–1013.
- 63.Felsenstein, J. Inferring phylogenies / J. Felsenstein. – Sunderland : Sinauer Associates, 1997.
- 64.Koonin, E.V. Orthologs, paralogs, and evolutionary genomics / E.V. Koonin // Annu. Rev. Genet. – 2005. – V.39. – P.309–338.
- 65.Yang, Z. PAML 4: Phylogenetic analysis by maximum likelihood / Z. Yang // Mol. Biol. Evol. – 2007. – V.24. – N 8. – 1586. – 1591.
- 66.Goldman, N.A. Codon-based Model of Nucleotide Substitution for Protein-coding DNA Sequences / N.Goldman and Z.Yang // Mol. Biol. Evol. – 1994. – V.11. – N 5. – 725. – 736.
- 67.Miyata, T. Two types of amino acid substitutions in protein evolution / T. Miyata, S.Miyazawa, and T.Yasunaga // J. Mol. Evol. – 1979. – V.12. – P.219–236.
- 68.Gillespie, J.H. Population genetics. A concise guide / J.H. Gillespie. – The Johns Hopkins University Press, 1998.
- 69.Rodi, D.J. DIVAA: analysis of amino acid diversity in multiple aligned protein sequences / D.J. Rodi, S. Mandava, L. Makowski // Bioinformatics. – 2004. – V.20. – P.3481–3489.
- 70.The R Project for Statistical Computing. – Режим доступа: <http://www.r-project.org>.
- 71.Mirny, M.A. Universally conserved positions in protein folds: reading evolutionary signals about stability, folding kinetics and function / M.A. Mirny, E.I. Shakhnovich // J. Mol. Biol. – 1999. – V. 291. – P.177–196.
- 72.Ичас, М. Биологический код. / М. Ичас. – М. : Мир, 1971.
- 73.Osawa, S. Evolution of the genetic code. / S. Osawa. – Oxford University Press, 1995.
- 74.Crick, F. The origin of the genetic code. // J. Mol. Biol. – 1968. – V.38. – P.367–379.
- 75.Freeland, S.J. Early fixation of an optimal genetic code / S.J. Freeland,

- R.D. Knight, L.F.Landweber, L.D.Hurst // *Mol. Biol. Evol.* – 2000. – V.17. – P.511–518.
- 76.Woese, C.R. The fundamental nature of the genetic code: prebiotic interactions between polynucleotides and polyamino acids or their derivatives / C.R.Woese // *Genetics.*– 1968. – V.59. – P.110–117.
- 77.Freeland, S.J. The genetic code is one in a million / S.J. Freeland, L.D. Hurst // *J. Mol. Evol.* – 1998. – V.47. – P.238–248.
- 78.Goldman, N. Futher results on error minimization in the genetic code / N. Goldman // *J. Mol. Evol.* – 1993. – V.37. – P.662–664.
- 79.Sella, G. The coevolution of genes and genetic codes: Crick’s frozen accident revisited / G. Sella, D.H. Ardell // *J. Mol. Evol.* – 2006. – V.63. – P.297–313.
- 80.Knight, R.D. Selection, history and chemistry: the three faces of the genetic code / R.D. Knight, S.J. Freeland and L.F. Landweber // *Trends in Biochemical Sciences.* – 1999. – V.24. – N 6. – P.241–247.